



Rising of Retracted Research Works and Challenges in Information Systems: Need New Features for Information Retrieval and Interactions

Peiling Wang

School of Information Sciences, The University of Tennessee, Knoxville

peilingw@utk.edu

ABSTRACT

This perspective paper analyzes the rising threat of retracted scientific works and the challenges of preventing the continued spreading and use of the retracted science; further, a framework is proposed for research and actions to effectively manage retractions in the information ecosystem. The precipitous increase in retractions of scientific publications is real and the complexity of retracting publications challenges current IR systems and people's information behaviors. Retracting published, especially peer-reviewed, papers in prestigious venues is a complex phenomenon involving various entities through often time-consuming processes. These publications may be accessible from the original venues, digital archives, or free-access databases, but these systems differ in retrievability and output. Many systems do not identify the retractions or reasons for retractions; most systems do not treat the retracted paper and its related notices (retraction or correction) as an integrated entity. Studies found that many retracted publications continue to be cited post-retraction as valid science. A new threat is the widely spread of retracted publications on social media. Retracting invalid scientific publications has serious implications in the real world. Based on current findings, we propose (1) a framework for further research; (2) a DOI resolution to integrate the documents related to retraction/correction; (3) a structured facet taxonomy for representing and indexing the retracted, corrected, or republished publications in databases; (4) a retraction registry or database with personalized AI helper for researchers to tract retracted publications; (5) an approach for understanding how retracted publications are circulated on social media.

CCS CONCEPTS

• **Information systems**; • **Information retrieval**; • **Retrieval tasks and goals**; • **Users and interactive retrieval**; • **Personalization**; • **Applied computing**; • **Document management and text processing**; • **Document management**; • **Document metadata**; • **Theory of computation**; • **Theory and algorithms for application domains**; • **Database theory**; • **Data integration**;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '23, March 19–23, 2023, Austin, TX, USA

© 2023 Association for Computing Machinery.

ACM ISBN 979-8-4007-0035-4/23/03...\$15.00

<https://doi.org/10.1145/3576840.3578281>

KEYWORDS

Retracted research papers, post-retraction retrieval, Context-sensitive tracking, New IR features

ACM Reference Format:

Peiling Wang. 2023. Rising of Retracted Research Works and Challenges in Information Systems: Need New Features for Information Retrieval and Interactions. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23)*, March 19–23, 2023, Austin, TX, USA. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3576840.3578281>

1 INTRODUCTION

A rapid growth in science was predicted 60 years ago by de Solla Price [1]. White (2021) reports that World publication output in science and engineering reached 2,940,807 articles for conference proceedings and peer-reviewed journals in 2020. Based on the data in Table SPBS-2 (<https://nces.nsf.gov/pubs/nsb20214/data#table-block>), the average annual growth was 5.15% for this century with the highest at 12.20% in 2005 [2]. Today, science is not only big but also open. The scientific publishing ecosystem is facing many new threats including predatory journals [3, 4], faked peer reviews [5], and paper mills that sell authors with fraudulent papers published in hijacked or reputable journals [6, 7]. Further, these illegitimate papers or journals are being cited by papers published in legitimate journals [8]. The precipitous increase in retractions in this century was observed by many studies e.g., [9], and was found more serious in medicine and life science [10, 11], but also occurred in computer science [12]. We searched three bibliographic databases, Web of Science (WoS), Scopus, and PubMed plus the retractiondatabase by the Retraction Watch to observe the retracted publications between 2001 and 2021 (Table 1). Distributions of the retracted publications were visualized to show the top categories and differences in classifications by Scopus and WoS (Figure 1).

This perspective paper will discuss the current research on retracted scientific research publications and the implications of the findings on researchers and medical information users in the real world. The main purpose of this paper is to raise awareness of the complexity of retracting published research works and advocate much-needed research to understand how information retrieval systems should help researchers and information users minimize the negative impact of retracted invalid science. The ultimate goal is to propose a model for AI-based personalized intelligent agents to effectively manage retractions and being-retracted works for researchers who collected and are using retracted publications in projects and publishing research results.

Table 1: Retracted publications in major databases (searches on December 29, 2022)

Pub. Year	Web of Science	Scopus	PubMed (retracted publication; express of concern; Corrected/republished)	Retractiondatabase (Retraction; Express of Concern; Correction)
2001	114		(150; ; 93)	(236; 15; 7)
2002	114		(170; 2; 80)	(297; 14; 15)
2003	143		(197; 2; 80)	(314; 20; 20)
2004	186	1	(251; ; 75)	(391; 21; 29)
2005	224	1	(309; 5; 76)	(477; 23; 23)
2006	307	3	(395; 3; 59)	(582; 26; 31)
2007	427	62	(398; ; 68)	(770; 21; 37)
2008	409	75	(453; 5; 43)	(752; 35; 36)
2009	750	753	(491; 4; 32)	(1590; 46; 59)
2010	870	4358	(556; 4; 47)	(5,240; 322; 69)
2011	1157 *	4121	(607; 3; 24)	(5,180; 69; 61)
2012	480	87	(672; 13; 26)	(1,322; 65; 78)
2013	441	312	(699; 16; 18)	(1,419; 68; 93)
2014	539	83	(926; 30; 28)	(1,563; 85; 77)
2015	582	119	(906; 29; 24)	(2,366; 81; 92)
2016	512	101	(764; 54; 13)	(2,047; 97; 72)
2017	525	106	(844; 54; 27)	(1,800; 114; 58)
2018	538	186	(1154; 69; 2)	(2,468; 180; 65)
2019	659	101	(1353; 136; 23)	(2,173; 148; 56)
2020	488	77	(1118; 282; 41)	(2,621; 137; 36)
2021	309	262	(725; 335; 28)	(2,548; 659; 18)
Total	9,774	10,808	(13,138; 1,045; 930)	(36,156; 2,246; 1,032)

* A publisher retracted a volume of 762 papers from Energy Procedia (DOI:10.1016/S1876-6102(14)00453-6)

Search results may have duplicated or missing records due to how IRs index the retracted documents and their retraction notices, as well as how search/filter functions are provided.

WoS query: (TI=(retraction OR retracted) AND DT=(Retracted Publication)) AND PY=(2001-2021)

Scopus query: TITLE (retracted OR retraction) AND (PUBYEAR > 2000 PUBYEAR < 2022) AND DOCTYPE (tb)

PubMed queries: “retracted publication”[pt], “expression of concern” [pt], “corrected and republished article” {pt} as three individual searches

Retractiondatabase.org: searched for publication year as a range (from MM/DD/YYYY to MM/DD/YYYY) with Nature of Notices limited to one at a time for Retraction, Expression of Concern, or Correction

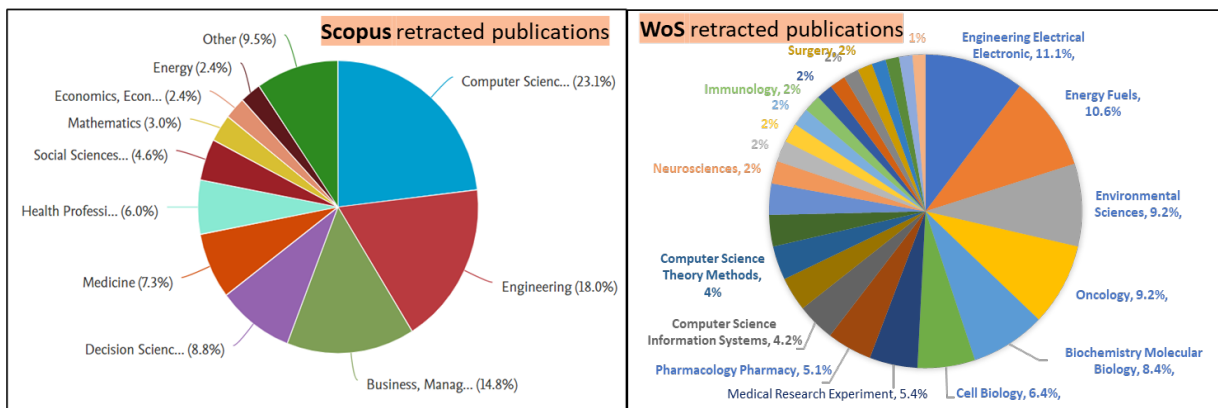


Figure 1: Subject categories of retracted publications from Scopus and WoS (Data in Table 1)

2 RETRACTION IN CONTEXT

The Committee on Publication Ethics (COPE) published The RETRACTION GUIDELINES [13] to advise editors and publishers on how to correct or retract erroneous or flawed publications, how to notify readers about a retraction of or an ongoing investigation of a publication, and what to do if an investigation results in inconclusive evidence to justify a retraction. The International Committee of Medical Journal Editors (ICMJE) also provides recommendations for corrections, retractions, republications, and version control referencing COPE's guidelines and flowcharts. ICMJE describes scientific misconduct in research as not limited to data fabrication and plagiarism. [14]

2.1 Defining Retraction and Retracted Documents

COPE's RETRACTION GUIDELINES state

"Retraction is a mechanism for correcting the literature and alerting readers to articles that contain such seriously flawed or erroneous content or data that their findings and conclusions cannot be relied upon. Unreliable content or data may result from honest errors, naïve mistakes, or research misconduct.

"The main purpose of retraction is to correct the literature and ensure its integrity rather than to punish the authors.

"Retractions may be used to alert readers to cases of redundant publication, plagiarism, peer review manipulation, reuse of material or data without authorisation, copyright infringement or some other legal issue (eg, libel, privacy, illegality), unethical research, and/or a failure to disclose a major competing interest that would have unduly influenced interpretations or recommendations." [13]

According to COPE, a *retraction notice* is for a single retracted article that should mention reasons and basis for retraction to inform readers, as well as identify by whom the article is retracted (e.g., by the authors or the journal). Partial retractions are not recommended because it is difficult to determine which parts may be relied upon. The *expression of concern* may be used as an alternative to retraction to alert the reader that the article is under investigation.

In further discussion, we use the terms documents, articles, papers, and publications interchangeably.

2.2 Current Practices of Retractions

The current practices vary by publishers, journals, and bibliographic databases regarding terminology, notices, and details. In Appendix A.1, we analyze a case to show the complexity of retracting a published paper. In this section, we provide examples for each type of post-publication status change.

2.2.1 Expression of concern as editorial notes. The data from PubMed and Retraction Watch (Table 1) show that Expression of Concern was used less than Retraction. The case below shows that the journal published an Expression of Concern when the editors were informed of data manipulation by one of the co-authors and

started the subsequent process of investigation. The follow-up Editorial Note informs "no further action" based on the experiments of two independent labs that largely confirmed the article's central conclusions. (See Appendix A for another example of an Expression of Concern with the outcome of retracting the article.)

1. The article was published in *Cell* on June 2, 2006, DOI: 10.1016/j.cell.2006.03.044
2. Expression of Concern was published on April 7, 2016. DOI: 10.1016/j.cell.2016.03.038
3. Editorial Note was published on September 8, 2016, regarding the outcome of the investigation. DOI: 10.1016/j.cell.2016.08.062

Although this case shows the journal's effort to ensure the accuracy of scientific results, the IR users may be confused because the three documents above each have a unique DOI and publication date, but identical title and authorship. Currently, the three documents are not integrated into a set for users. Integrating the three documents is important for understanding the history surrounding the concerns.

2.2.2 Retraction of published articles. The data from PubMed and Retraction Watch (Table 1) show that Retraction is the most among the three types of status change. This article in *Nature* was retracted in less than one year by the authors.

1. The article was published in *Nature* (26 May 2021) DOI: 10.1038/s41586-021-03503-5
2. Retraction Notice was published on 24 February 2022 by the authors, DOI: 10.1038/s41586-022-04525-3

From the original article, two sentences under the title section: "This article was retracted on 24 February 2022. This article has been updated." Each of the sentences was link is to the same Retraction Notice, which states:

"..... We made an error in calculating the global mean precipitation: we used arithmetic averaging to calculate the mean, instead of calculating a spatially weighted mean to account for the changing grid box size with latitude. As a result, the magnitudes of the global mean precipitation time series were underestimated. This impacted the subsequent calculation of global mean evapotranspiration, resulting in the mean evapotranspiration values being underestimated and altering some results. We are therefore retracting this article. We thank Ning Ma and others for bringing this error to our attention."

As above, there is no mention of a new article or the updated calculation results. The article's webpage has "RETRACTED ARTICLE:" preceding the title and its [Cite this article](#) provided citation also has "RETRACTED ARTICLE:" preceding the title. In ResearchGate, however, the title of this article's record does not have the word "Retracted," but the preview of the first page has the stamp "RETRACTED ARTICLE." In a controversial site for free pdf articles, Sci-Hub, we were able to download the pdf file using DOI. The version was not stamped "RETRACTED ARTICLE" (likely the version before the retraction).

As of December 25, 2022, a search of Web of Science using DOI found the article marked as "Highly Cited Papers" and "Hot Papers,"

but the record does not have a “Retracted Article” leading title, nor a link to the Retraction Notice. A search of the title will retrieve both the record for the article and the record for its RETRACTION Notice. The same search of the DOI in Scopus resulted in the article record with an Update notice below the title linked to the Retraction Notice, but the article’s record does not have “Retracted Article” in the title. PubMed also has two mutually linked records for the article and its Retraction Notice. Although the autogenerated citation by PubMed added “Retraction in:”, the term “Retracted Article” was not leading the title. All three bibliographic databases did not add “Retracted Article” in their records’ titles. We found that in the record from SAO/NASA Astrophysics Data System (ADS), the title has “Retracted Article:” at the beginning. Retractiondatabase has one record that integrated the Retraction Notice with the article.

2.2.3 Correction (updated, erratum, corrigendum). The data from PubMed and Retraction Watch (Table 1) show that Correction occurred the least. This category is important because the corrected part may or may not change conclusions or other findings on which users may rely. The following two examples show different ways by which journals correct errors. The first example is simply to correct an inadvertently neglected author in authorship. The author as the 7th (of the 16 authors) was added to both HTML and pdf (no different versions).

1. The article was published in *Cell* on February 25, 2016. DOI: 10.1016/j.cell.2016.02.006

2. CORRECTION was published on April 07, 2016. DOI: 10.1016/j.cell.2016.03.037

The second example published a correction more than 10 years later.

1. The article was published in *Proceedings of the National Academy of Sciences* on January 28, 2008. DOI: 10.1073/pnas.0707594105

2. Correction was published by the journal online on March 19, 2018. DOI: 10.1073/pnas.1803343115

In this case, the investigation was conclusive:

“ The Committee concluded that the manipulation of the images in Fig2D could only have occurred intentionally, representing instances of scientific misconduct. The Committee could not definitively attribute the research misconduct to any individual.

“One author then was able to identify an image copy of the original The correction does not impact interpretation of the data.

The corrected figure has been evaluated by the editor and approved for publication.”

Although the article’s webpage has a note: “THIS ARTICLE HAS BEEN CORRECTED +” with a link to the Correction, the pdf version of the original article still has the incorrect image. PubMed used “Erratum” instead of “Correction” to include the corrected image. The article is accessible in ResearchGate with all the Figures loaded by one of the co-authors, including the incorrect image. We have sent a message to this author and the link to the Correction notice in ResearchGate. This article has been cited 978 times in WoS all databases)

2.2.4 Retracting and republishing a paper. This type is not easy to track from databases. Below are three different situations when a

paper was republished. The first case is about retracting a paper and republishing it after revisions; the retracted paper was removed from the journal.

1. The article was published in *Population Health Management* online on 28 Nov 2018. DOI: 10.1089/pop.2018.0024, which has “Retracted:” leading the title, but the page is the same notice as Retraction below

2. Retraction was published again online on 27 Mar 2019. DOI: 10.1089/pop.2018.0024.retract

3. Republication under the same title and authorship was published on 26 Sep 2019. DOI: 10.1089/pop.2019.0054

The retraction was triggered by the developer of a cited instrument who had requested changes to the article. The authors decided to revise the paper by removing the instrument. The case is beyond typical errors or flaws as reasons to retract (see 2.1). The retracted article’s webpage provides a paywall (\$51 for 24 hours) access to the full text, but if the notice is to remove the retracted version. The reason for the retraction explained in the notice only makes it necessary to compare the two versions, despite the claim:

“The retraction serves to remove the published version of the article that contains the MMAS-8 scale and a revised version that does not contain the tool or any references to it will be published. The elimination of the scale does not alter the results or conclusions of the study.”

There was no link to the republished paper from the retracted article page or the Retraction page. We, however, found the original article in Sci-Hub. In a cursory comparison, the republished article is one page shorter and 9 references fewer than the retracted article. Because the retractiondatabase does not index “republishing,” only the Retraction notice and the retracted article as one record.

The second case is about retracting a paper and republishing it with the original paper attached to the new version.

1. The article was published in *eBioMedicine* on August 25, 2020. DOI: 10.1016/j.ebiom.2020.102957

2. The article was republished on March 20, 2022. DOI: 10.1016/j.ebiom.2022.103923

3. Retraction and republication notice by The Editors on April 4, 2022, DOI: 10.1016/j.ebiom.2022.103966

The retraction and republication notice state:

“ On November 17, 2021, the authors drew our attention to an error the group allocations to the magnesium sulphate and placebo study groups in the paper were inadvertently reversed, making the findings, as reported, unreliable.

“Today we retract the previous version and republish online the corrected version of the article DOI:10.1016/j.ebiom.2022.103923 in which the results are unchanged apart from the treatment group assignment. The figures and tables have been corrected accordingly, and the overall message has been modified

“The previous version of the Article has been added to the appendix in the new version and is marked retracted.”

The Retraction Notice provided links to both the retracted article and the republished article. The republished article included the retracted version, which also helps users to compare the differences and evaluate the errors.

The third case is about retracting from one journal and republishing in another journal.

1. The article was published in *Isr J Health Policy Res.* 2019 May. DOI: 10.1186/s13584-019-0314-8

2. Retraction of publication note in *Isr J Health Policy Res.* 2019 July. DOI: 10.1186/s13584-019-0330-8

3. Republished in *BMC Health Serv Res.* 2019 Jul 3;19(1):445. DOI: 10.1186/s12913-019-4268-x

Both journals are part of Springer Nature, thus a quick turnaround for the retraction and republication. The Publisher states that the article was published in *Isr J Health Policy Res* in error and is republished in *BMC Health Serv Res*. The republished paper is identical in contents and page layout, except for the journal. The republished paper does not have any notes about replacing the previously published and retracted paper.

The retracted paper is still accessible in *Isr J Health Policy Res* with the term “RETRACTED ARTICLE” on all the pages. Other bibliographic databases such as academia.edu and ResearchGate.net also have the original retracted paper’s pdf file that is not marked “RETRACTED ARTICLE” on pages. The record in ResearchGate has “Retracted” above the title, but the record in Academia does not indicate the paper’s Retraction status.

2.2.5 Accessibility of retracted or corrected articles and their notices. Although some journals provide free access to retracted or corrected articles, and their notices, most journals we checked still have a paywall for the retracted or corrected articles or their retraction or correction notices. When a journal provides free access to retracted or corrected articles at some point, the previously retracted or corrected papers and their notices may still be behind a paywall.

The retracted *Nature* article in 2.2.2 is still subscription or pay-for-access although its Retraction Notice is free. The following examples show different ways of handling access by *Science*:

1. The article was published in *Science* on 16 July 2010 (online on 3 June 2010). DOI: 10.1126/science.1188888

2. Correction notice was dated 9 September 2011 in section Corrections and Clarifications (Figure 2, Appendix B)

3. The Correction notice was added to the end of the paper’s pdf version

Access to the Correction notice published in Corrections and Clarifications is not free (Figure 2)

Although the one-page section Corrections and Clarifications is not free, the corrected article is accessible to registered community users (see below)

We tracked another *Science* paper that was published in 2006 and corrected in 2016.

1. The paper was published in *Science* on 21 Apr 2006. DOI:10.1126/science.1126088

2. Erratum was published on 22 Jan 2016. DOI: 10.1126/science.aae0382

For this article, the Erratum is free, which noted that a figure in the original paper was incorrectly assembled. A corrected figure was included in the Erratum and the original conclusion remained valid.

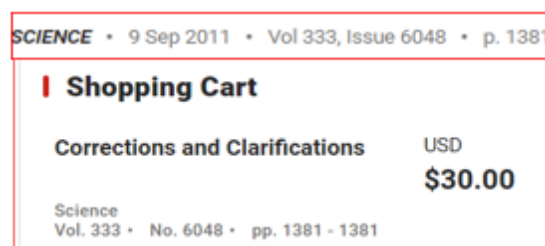


Figure 2: Paywall for Corrections.

However, the Erratum did not provide a link to the original paper. We searched and found the original article at <https://www.science.org/doi/10.1126/science.1126088>, which has a note on Erratum and a link to Erratum. The article is also free to registered community users.

“REGISTER FOR FREE TO READ THIS ARTICLE

As a service to the community, this article is available for free. Login or register for free to read this article.”

We checked some recently retracted articles in *Science* and most of them are free to registered users, such as the 2009 *Science* paper retracted in 20110 (DOI:10.1126/science.1174094). However, a 2004 *Science* paper retracted in 2009 (DOI:10.1126/science.1089509) is still behind a paywall at this writing. We, however, could not verify when or which retracted or corrected articles are free to access.

2.3 Retrieval of Retracted Publications in Databases

The four databases we searched for retracted articles (Table 1) provided different search features and results to identify and track retracted publications. Retractiondatabase is the most inclusive and easy-to-use specialized retraction database. Subject terms can be selected from a hierarchical classification system. The field Nature of Notice defines 4 types: (1) Retraction, (2) Expression of Concern, (3) Correction, and (4) Reinstatement (we did not examine reinstatement as it deals with an online publication temporarily not accessible). Searches can limit to two types of date ranges, for publications and/or retractions. The form-fill interface is intuitive and efficient to search for retracted authors, known titles, and DOIs. For more comprehensive and open-ended searches producing a larger number of results, such as reasons for retractions, subjects, country, journal, publisher, and dates, only 50 records are displayed. For example, the search for “United States” in Country AND “Retraction” in Nature of Notice found 4,604 items, but only 50 were displayed. The only way to get the 4,604 items is to narrow the search by date ranges to produce 50 or fewer items repetitively, which is possible but inefficient. Users may also wish for functions such as exporting records in results. When we collected data for Table 1 (last column), we only needed the number of items by year, not the actual items.

PubMed is a specialized database for biomedical literature with an advanced query builder to select fields with Boolean operators. The Publication Type field has the index terms “corrected and republished article,” “expression of concern,” “published erratum,” “retracted publication,” and “retraction of publication.” There are

126,574 records indexed by “published erratum” including both correction notices, erratum, corrigendum, author response, etc. Each correction notice is a record in PubMed, thus multiple corrections were not integrated into the article’s record. For example, the three records are cross-referenced in PubMed.

1. The article was published in PLoS One, on Jun 17, 2014. DOI: 10.1371/journal.pone.0099868

2. Correction was published on Sep 2, 2015. DOI: 10.1371/journal.pone.0137930 (to correct a Table)

3. Correction was published on Jan 11, 2018. DOI: 10.1371/journal.pone.0191394 (to correct 2 Tables)

Searching for known publications (using author, title, or DOI) can easily find the retracted or corrected publications or their notices of status changes. Advanced query builders with guided indexes can perform complex searches,

Web of Science (WoS) uses the filtering field Document Type to assign index terms: “Correction”; “Correction, Addition”; “Expression of Concern”; “Retracted Publication”; “Retraction” and “Withdrawn Publication” to the articles or their notices. In WoS, advanced searches require terms to be entered in one of the required fields such as Author, Title, etc. before filtering by Document Type. Because Document Type is not a search field, we were unable to break down our search results in Table 1 similar to PubMed and retractiondatabase.org. When we checked records for the case in Appendix A, we found the article’s 2014 Expression of concern was indexed as “Editorial Material” for Document Type. It was unclear when WoS introduced the above index terms.

Scopus, on the other hand, also uses Document Type to handle changes to publications. Currently, “Erratum” and “Retracted” are the only two terms used to index corrected or retracted publications. “Note” or “Editorial” is used for the retraction notice, expression of concern, etc.

Our testing searches show that WoS and Scopus do not retrieve retraction-related records easily or display them as a related set. Reasons for retractions or corrections are not indexed. PubMed is a well-designed and implemented database for searching biomedical and health-related articles and their retractions, but the index term “published erratum” is not effective in narrow search results.

Last but not least, the only database that integrates the retracted publication and its notices as a single entity is retractiondatabase.org, which is important for searchers to get all related records for the retracted document. WoS, Scopus, PubMed, and ADS (see example in 2.2.2) index the retracted publication and its notices as individual records. Therefore, depending on how searches are formulated and reiterated, users can miss these related documents.

2.4 Summary

Our investigations intended to provide examples to corroborate reported studies on issues and draw attention to challenges facing retractions in current information ecosystems. Retraction is a complex phenomenon with many moving pieces over time, which makes it difficult for the current information systems to handle. While COPE’s RETRACTION GUIDELINES provide general principles, the specifics for the implementations are decided by the publishers or journals. To summarize our observations:

1. The word *retraction* cannot define all types of changes that occurred post-publication
2. The process of retractions can span decades, which made it difficult to correct scientific literature
3. There are many parties involved in the process
4. Greater transparencies are needed for retraction notices to provide details to preserve useful parts of the work and help users to understand the nature of the problems
5. The notices should include, but often lack, the reasons, history of the process, event-evoked investigation, etc.
6. A standardized vocabulary is needed for describing and indexing the retracted works by journals and databases
7. Inconsistency in marking the retracted articles affects the use and citing
8. One retraction notice should be for one retracted article even if multiple publications of an author were retracted once (e.g., DOI: 10.1097/SHK.0000000000001707 retracted three articles in one Notice)
9. Retracted documents and their notices must be clearly linked in both directions and output as an integrated entity
10. Paywalls for retraction notices or retracted works make it difficult to correct errors or use valid results (e.g., Figure 2; to access the one-page Corrections of an article published in *Thyroid* DOI: 10.1089/thy.2010.0376.cxn, \$51 for a 24-hour online view)

3 IMPACT OF RETRACTED SCIENCE

3.1 On Medical Practice

Flawed or erroneous scientific publications have a direct impact on research. As an example, the retractions of published fraudulent studies led by Boldt, a highly prolific author, were investigated for research misconduct and data fabrication. On March 12, 2011, the Editors-in-Chief of 18 journals jointly made a statement to retract 88 articles out of 102 published by Joachim Boldt for unverified IRB approval [15]. Following this first set of retractions, more of Boldt’s publications were retracted over time; Retraction Watch’s database has 169 retracted articles by Boldt to date including a 1990 article retracted on 10/13/2022 [16].

To assess the impact of the retractions of Boldt’s publications, a meta-analysis and systematic review published in 2013 concluded that excluding his 7 clinical trials involving 590 patients, “hydroxyethyl starch was found to be associated with increased mortality among 10,290 patients.” [17] The clinical guidelines cited in Boldt’s studies were also reviewed and replaced by new guidelines but changing the practice of every hospital is a much harder endeavor.

In an investigation of the impact of 12 retracted trial reports by different researchers, the study [18] found that five of the nine clinical guidelines cited in these trials would likely be altered if these retracted trials were excluded. A large-scale study evaluated 788 retracted papers with a focus on 180 retracted primary studies involving humans; the primary studies treated 9,189 patients. The 851 secondary studies that cited a retracted primary paper treated 70,501 patients. Papers retracted for fraud treated more patients per study than papers retracted for error. [19]

Additional examples of a high number of retractions in perioperative medicine spoke the increasing concern as the phenomenon

impacts the trust between authors and professionals, and relationships between doctors and patients. Although these retracted fraudulent publications were by a small number of researchers, future recurrences need to be prevented. [20]

3.2 On Research

In addition to the impact of unreliable research on medical guidelines and patient care, as shown in the examples in 3.1, research built on fraudulent publications is at risk of retraction (The Domino effect). Some articles, although recommended by domain experts, were later retracted [21]. More seriously, continued citing retracted invalid publications are “likely to spread misinformation from the retracted paper.” [22] In a survey of the authors who cited retracted trials, the researchers found that a total of 45 of 88 citing publications had findings likely to change if the retracted trials were removed; of the 45, 39 were likely substantially impacted. [22] When the retracted papers were cited as valid science whether pre- or post-retraction, the citing papers inadvertently spread invalid findings. Several studies tried to understand how the retracted papers were cited in context. One study [23] found that the majority of the 238 citations of the retracted articles were positive even though the retractions were due to ethical misconduct, data fabrication, and false reports. Another study [24] reported that only 5.4% of the 722 citations acknowledged the retractions and the retracted papers were mostly cited as related work. Analysis of 961 post-retraction citing articles of 77 retracted articles found that 86% cited without mentioning retraction or flaws; only 2% debunked the cited articles without mentioning retractions [25].

3.3 Spreading of Retracted Publications on Social Media

In today’s social-technical environment, social media such as Twitter, blogs, Facebook, digital news outlets, etc., are circulating retracted publications on a much larger scale and faster [26–29]. Researchers found that the subsequent retraction is not an effective tool to reduce online attention to problematic papers. [29] The topic is important but scarce literature exists as research in this area has just started.

3.4 Post-retraction Citing

The time taken to retract papers varied widely. The longest at this point is a 1975 paper that was retracted in 2020 by Editors for being unable to verify parental consent for reported cases (PMID:1233443). To observe post-retraction citing, studies typically counted citations one or two years after the retraction as post-retraction citations. Given the citation window is approximately 5 years, retracting an old paper may not result in high post-retraction citations. However, highly cited retracted papers seemed to show a longer citation lifetime. [21, 30]

Retracted publications may still have valid parts that can be cited. The problem is that many authors are citing retracted papers without acknowledging the retractions or pointing out the problems in the papers. [25] Several reasons may contribute to this phenomenon. Some authors were citing from the citations in published papers instead of searching the databases where retractions were clearly indicated [31]. Due to the time lag between the published

original paper and its retraction notice, researchers might have collected the papers in their pre-retraction versions without tracking their status again during writing and publishing research outcomes. Inconsistent retraction practices across journals vary from informative and transparent to deeply obscure or lacking clarity. Because retraction processes and notifications vary by publishers and journals, measures such as “retraction check” are necessary to avoid citing retracted literature. [32] In fact, a “retraction check” will require many more strategies than typical known-item searches as illustrated above. The issues summarized in 2.4 are likely also contributing to post-retraction citing.

3.5 On Users’ Information-seeking Behavior

To prevent using retracted scientific papers, Enago Academy suggests: “You can now easily weed out bad science by searching for retracted papers on Retraction Watch.” [33] As shown in section 2.3, we can rank retractiondatabase.org as a top choice for finding articles that were retracted, corrected, or concerned. PubMed should be searched for as well for biomedical and health-related publications. WoS provides more data such as funding sources and citing documents.

With the many other options for users to discover information, retracted papers are stored and searched differently across information platforms. Searches need to be conducted cross-databases using different strategies and tactics. For example, knowing an author’s work has been retracted, continued tracking of the author’s other publications becomes necessary. We searched for retracted articles by Joachim Boldt (see also 3.1): 40 records in PubMed using query (“retracted publication” [Pt]) AND (Joachim Boldt[Author]); 136 records in Scopus including Erratum (93), Retracted (26), Editorial (9), and Note (8); 35 records in WoS for document types as Correction (16), Retraction (12), and Retracted Publication (7); 169 records in retractiondatabase.org, which integrated retracted articles with their notices in one record.

The wide-ranged differences in results from these databases are due to how the databases indexed the documents and what query features are provided. Users will need to adapt to each system to identify or track retracted or corrected publications across different databases. Because a published paper can be retracted or corrected anytime to avoid using or citing invalid results in retracted publications, researchers will need to monitor the literature they use.

Last but not least, users of Sci-Hub downloading full-text pdf files need to check if the files were the pre-retraction version by searching both PubMed and retractiondatabase.org.

4 FACE THE CHALLENGES OF RETRACTIONS

It was pointed out more than 30 years ago that:

“Methods currently in place to remove invalid literature from use appear to be grossly inadequate. Regardless of strides made in controlling fraud, error is generally considered an inherent and inevitable aspect of research, and efficient removal of invalid information from the literature would serve science well.” [34]

As pointed out in a recent synthetic discussion, the situation has not improved much:

“... current models to correct the literature are unable to effectively deal with corrections of errata or retraction errors when errors are found in these correction notices. Even though the process for the detection and correction of error and fraud might be fairly well established and “standardized”, such as in COPE or ICMJE guidelines, inter-journal and inter-publisher variability, including editorial responsibilities, will continue to limit the effective correction of erroneous and fraudulent literature globally.” [35]

There is an increased awareness of the phenomenon of rising retractions in scientific publications. The RISRS project [36] made four recommendations:

1. Develop a systematic cross-industry approach to ensure the public availability of consistent, standardized, interoperable, and timely information about retractions
2. Recommend a taxonomy of retraction categories/classifications and corresponding retraction metadata that can be adopted by all stakeholders. (COPE and NISO are working on this taxonomy.)
3. Develop best practices for coordinating the retraction process to enable timely, fair, unbiased outcomes
4. Educate stakeholders about publication correction processes including retraction and pre- and post-publication stewardship of the scholarly record

Calls also from scientists for more transparency in retraction notices [37, 38] and a self-retraction system for honest errors [39, 40]. Retraction notices have been analyzed for not following COPE Retraction Guidelines [41, 42]. Their findings include:

1. Retraction notices lack or are highly inconsistent within journals and across publishers
2. 53% of retraction notices are made without identifying who initiated the retraction; only 15% were initiated by the authors
3. Retracted articles were not clearly marked as being retracted
4. Reasons for retractions were ambiguous
5. Extents of retraction (e.g., partial retraction vs. erratum)

The inconsistencies in how publishers and databases are handling retraction notices and retracted articles make the retracted publications difficult to identify or track. Researchers suggest that

“...by making the retraction information more clearly discoverable and standardized, the number of unintentional and unacknowledged citations of retracted literature will be reduced.” [42]

5 CALL FOR RESEARCH AND ACTIONS

Documents in information systems include attributes and access points such as authorship, title, journal, conference, subject, classification, publication date, etc.; information retrieval systems (IRs) store documents or their representations to enable searching and filtering during interactive information retrieval (IIR). Most IRs can follow up with users with personalized alerts on new content based on searches or stored information needs. However, the current bibliographic databases are not designed to manage the complexity of retracted or corrected documents. Users need to track publications they have collected thus they need more than searching and interactions with the systems.

If a document can be subsequently retracted or corrected, the concept of a document is no longer a permanent product, and its representations and indexes also need to change along with the status. Most importantly, the link between a retraction, or correction notice as a document and its retracted or corrected document should be in both directions; in other words, the retraction or correction notice of the article is an integral part of the retracted document. The cases of expression of concern about a publication bring challenges for both IRs and users. First, the notice may lack sufficient information on the causes. Second, it is uncertain in both time and outcome of the investigation.

5.1 A Framework for Interactions between Research and Retracted Publications

The factors related to interactions between research use of literature in context and retractions of invalid science in processes are identified in Figure 3. A typical research project may be divided into four main stages:

1. S_1 starts the project during which relevant publications are collected
2. S_2 works in progress during which information seeking is less active
3. S_3 produces findings when researchers search for new publications but rarely re-track already used publications
4. S_4 publishes research outcomes when manuscripts are submitted, revised, and published

Retractions of publications may take three different processes:

1. P_1 was initiated by the author and the editor issued a retraction notice
2. P_2 a process started an investigation by the authors' affiliations or interinstitutional review board (e.g., the case established → journal published Expression of Concern → the decision)
3. P_3 made by the journal to retract an article for reasons, such as duplicated publications, publications having cited retracted publications or additional publications of a fraud author

Each of the research stages (S_n) may intersect with the process of a published paper being retracted or corrected (P_m). For example, if the project is at S_1 the published paper D_1 is retrieved and useful for the project at hand. As the project continues, one of the changes to D_1 may happen in these processes:

1. P_1 : If the author requested to correct or withdraw the paper D_1 , the journal will publish a correction D_2 or a retraction notice D_4
2. P_2 : Journal editors retracted D_1 and published a retraction notice D_4
3. P_3 : Article D_1 is under investigation, the journal publishes an expression of concern D_3

P_3 is the most complicated process. The decisions will depend on the outcomes of the investigation to justify a correction D_2 , a retraction D_4 , or a republication D_5 ; if the investigation is inconclusive, there will not be further action, thus an editorial note concludes the case.

The inseparable relationship of these documents during the process must be integrated as an entity beyond simply cross-referencing. That is, if any of the documents about the same work (the retracted article, its retraction notice, or republished article)

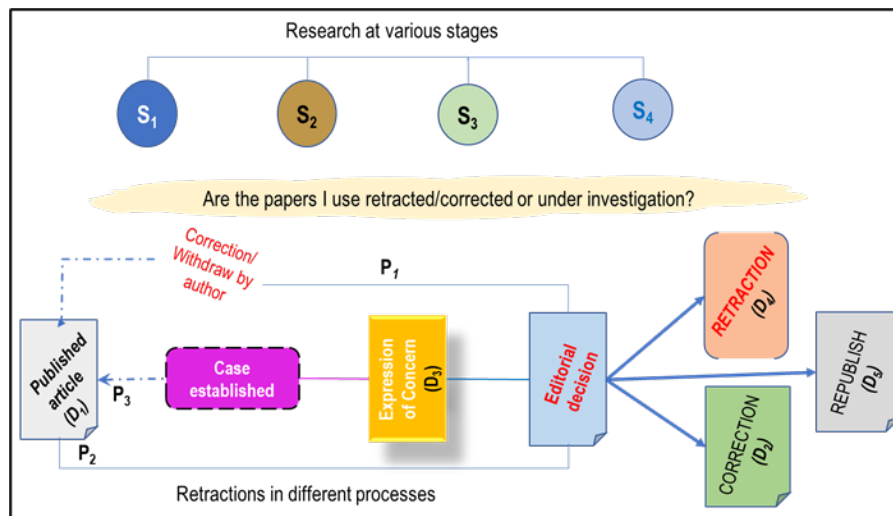


Figure 3: Framework for research use of retracted publications

are in the search results, the other related documents should automatically be included as a set for the searcher.

What about the searchers who retrieved the document before its retraction? We will need effective ways to reach them regarding the changed status of the papers which they have collected and are using. The description of the retracted documents will need new attributes that do not currently exist in IR models or databases. These attributes include the history of the events in the process, specific changes, reasons, etc. are useful information for users. Currently, as a compromise, PubMed and WoS use the filed document type to index retracted publications and their notices. As the number of retractions and corrections increases as an inevitable phenomenon, finding and tracking retracted publications are more problematic and the observed inadvertent use of invalid scientific publications will also continue.

5.2 Need DOI Resolution for the Set of Documents

The current DOI mechanism is insufficient to track the full history and background of the retraction and correction or republication over time. A more robust method is needed, such as Double-DOI [43]. Another approach is to establish a retraction/correction registry to enable access to the whole set of documents related to the process. That is if a searcher only knows one document in the set (e.g., retraction notice, retracted article, republished article, and correction notice), the results will provide the set of documents with history. The registry must use standard taxonomy (see below) to differentiate types of publication changes and enable searches for reasons.

Other types of relationships among documents such as Letters to the editor regarding a published article and Response by the author also require integrated output. For example,

1. The original article DOI:10.1089/thy.2010.1654
2. Letter to the editor DOI: 10.1089/thy.2011.0179
3. Response by the author DOI: 10.1089/thy.2011.0179.rs

5.3 Need a Taxonomy to Standardize Indexing and Searching

The use of a single word “retraction” to tag different types of changes (i.e., retracted, corrected, concerned, or republished) is inappropriate because a paper with errors may still be useful (i.e., “the retraction cause did not invalidate the study results.” [44]) Thus, information seekers need to have sufficient information to make decisions on the use of such publications. In a 2011 News Feature article published in *Nature*, the suggestion by Rennie, then editor of JAMA, was quoted as “reserving the retraction mechanism exclusively for misconduct” and the suggestion by Nicholas Steneck, Founder of World Conferences on Research Integrity, was quoted as “A better vocabulary for talking about retraction is needed. . . . Also useful would be a database for classifying retractions.” [44] A taxonomy should standardize vocabulary for retraction/correction practices and help information seekers find and identify publications effectively.

What needs to be in the Taxonomy? We recommend adopting structured facet categories to include STATUS—REASON—INITIATOR—EXTENT—HISTORY

1. STATUS
 - 1.1 Withdrawal
 - 1.2 Retracted
 - 1.3 Corrected
 - 1.4 Republish
 - 1.5 Expression of concern
2. REASON
 - 2.1 Duplicated publication
 - 2.2 Error
 - 2.2.1 Data
 - 2.2.2 Statistical
 - 2.3 Findings
 - 2.3.1 Interpretation
 - 2.3.2 Unsupported
 - 2.4 Presentation

- 2.4.1 Figure
- 2.4.2 Table
- 2.5 Misconduct (intentional act)
 - 2.5.1 Fabricated data
 - 2.5.2 Fraud
 - 2.5.3 Plagiarism
- 2.6 IRB violation
- 3. INITIATOR
 - 3.1 Author
 - 3.2 Institution
 - 3.3 Funder
 - 3.4 Editor (Journal)
 - 3.5 Publisher
- 4. EXTENT
 - 4.1 Full
 - 4.2 Partial
 - 4.2.1 Findings valid
 - 4.2.2 [need more cases to decide]
- 5. HISTORY (a composite multi-valued attribute)
 - 5.1 Date
 - 5.2 Status [choose from 1]

The following examples show how the faceted taxonomy may be used to classify or tag the articles:

1. Retraction—IRB violation—Institution—Full—{date1, Expression of concern; date2, Retraction}
2. Correction—Error, Statistical—Author—Table—{data1, Corrected}

The facet HISTORY will be a composite attribute in the database allowing multi-values. A free text field should also be used to provide details or specifics.

A complete taxonomy should be built using a bottom-up approach based on what has been retracted. Retraction Watch has a list of reasons. One study [38] suggested 17 reasons; another study [45] derived 13 reasons from the retracted articles collected in BioMed Central.

5.4 A Global Registry for Retractions and Corrections of Publications

One of the best solutions might be to design and develop a global registry for tracking retractions (and corrections) of publications. The registry will integrate the set of documents and index/tag them using the standardized taxonomy. For these documents to be of value, information should focus on transparency and helping users evaluate the work for adequate use or critique. This registry should also be a platform for authors of these documents to make contributions of reflections or share learned lessons.

This database model will be developed as a research project based on the framework.

5.5 Understanding How Retracted Publications are Circulated on Social Media

Social media platforms as open and dynamic systems provide new channels to share publications including retracted papers, especially on Twitter. These channels influence scholars, practitioners, journalists, and the public. [26–29] The study [27] observed that the retracted papers were disseminated widely through multiple

channels months after publication and sooner than other papers; retractions had a limited impact in reducing the spread of such dissemination. The phenomenon raises a new challenge: how to handle retractions of scientific publications across various social media platforms? Each platform has its unique structure and functions as well as its adopters. Studies of Tweet spreading retracted publications have moved beyond the counting of the tweets and retweets over time to analyzing the text to identify if critical tweets on circulated papers can provide some discussion of or clue to problematic papers [27, 46].

One way to conceptualize information interactions in a social network system is to identify the subnetworks in which specific retracted papers were circulated and discussed. In a substructure of a social media network, members may belong to different subgroups and interact on different topics. The nodes represent members and ties represent interactions that may happen from the influential member to others or between members on issues and topics. Using the concept of *opinion leader* from the diffusion of innovation theory by Rogers [47] and the concept of *centrality* and *temporal moves* in social network analytics, effective methods can be developed to impact these subnetwork groups in which retraction notices should immediately follow the previous circulated retracted publications.

6 CONCLUDING REMARKS

The effort in this research identified the many challenges of the rising retractions in today's scientific publications points with a focus on peer-reviewed publications. It must be pointed out that the preprints repositories (such as arXiv) or post-publication open peer-peer systems (such as F1000Research) are facing similar if not more challenges, which this paper has not been able to address. The current mechanisms and practices for handling retractions and corrections of scientific publications by publishers, journals, and the IR systems are diverse and inconsistent. Information science as an interdisciplinary field can contribute to more effective solutions, some of which information retrieval and interaction can play significant roles:

1. The Retraction Guidelines should be revised to be inclusive and increase specificity: Guidelines for Retraction and Correction
2. A dedicated information system/database or global registry for publication retraction/correction should be designed and implemented for information seekers to get all questions answered about a publication whose status changed (currently a user must access different systems to track retracted publications); the system can also be a platform for authors of these publications to share reflections and lessons
3. A new data model will need to integrate the set of documents as one entity (including the original publication, status change notices, corrected version, and republished version)
4. The representation of this entity needs to use standard taxonomy to index/tag what type of changes were made to the publications, the reasons, the initiator, and if any parts are still valid after the change
5. The registry and databases should have an AI-based personalized intelligent agent for reaching out to researchers who collected and are using these publications

6. In the era of open science, consider the opt-in option for authors to publish their communications with the editors or explanations about the issues as part of the entity to help inform the scientific community
7. Research needs to understand how publications are being circulated on social media (who posted them) and the interactions of the subnetworks; based on this knowledge, we can develop strategies and design solutions to ensure retractions reach these subnetwork groups promptly
8. Citing retracted publications need to identify valid results and problematic aspects to minimize the continued spreading of retracted science (for this, citation guidelines and online generated citations must mark retracted or corrected publications)
9. Information behavior solutions: beyond searching to find relevant documents, researchers need to be aware of the changing status of the documents that they collected and used, especially if they are in fields with high retractions
10. Fundings are needed for research and the development of new models with new features for databases storing and retrieving retracted publications.

ACKNOWLEDGMENTS

Special thanks go to Emrie Gibert for editorial assistance.

REFERENCES

- [1] Derek J. de Solla Price. 1963. *Little Science, Big Science*. Columbia University Press, New York.
- [2] Karen White. 2021. Publications Output: U.S. Trends and International Comparisons. Science and Engineering Indicators. (October 2021). Retrieved from <https://nces.nsf.gov/pubs/nsb20214/data#supplemental-tables>.
- [3] Amanda Ross-White, Christina M. Godfrey, Kimberley A. Sears, and Rosemary Wilson. 2019. Predatory publications in evidence syntheses. *Journal of the Medical Library Association* 107, 1 (Jan. 2019), 57–61. DOI: 10.5195/jmla.2019.491
- [4] Agnes Grudniewicz, David Moher, Kelly D. Cobey, et al. 2019. Predatory Journals: No Definition, No Defense. *Nature* 576 (Dec. 2019), 210–212.
- [5] Xingshun Qi, Han Deng, and Xiaozhong Guo. 2017. Characteristics of retractions related to faked peer reviews: an overview. *Postgraduate Medical Journal* 93 (Aug. 2017), 499–503.
- [6] Horacio Rivera and Jaime A. Teixeira da Silva. 2021. Retractions, Fake Reviews, and Paper Mills. *Journal of Korean Medical Science* 36, 24 (Jun. 2021), 1–7. DOI: <https://doi.org/10.3346/jkms.2021.36.e165>
- [7] Anna Abalkina. 2022. Publication and collaboration anomalies in academic papers originating from a Russian-based paper mill. The 9th Peer Review Congress: *Enhance the quality and credibility of science*, September 8–10, 2022, Chicago, IL, 24.
- [8] Anna Abalkina, Guillaume Cabanac, Cyril Labbé, and Alexander Magazinov. 2022. Improper Legitimization of Hijacked Journals Through Citations. The 9th Peer Review Congress: *Enhance the quality and credibility of science*, September 8–10, 2022, Chicago, IL, 75–76.
- [9] Li Tang, Guangyuan Hu, Yang Sui, Yuhan Yang, and Cong Cao. 2020. Retraction: the “Other Face” of research collaboration?. *Science and Engineering Ethics* 26 (Mar. 2020), 1681–1708. doi:10.1007/s11948-020-00209-1
- [10] Michael L. Grieneisen and Minghua Zhang. 2012. A comprehensive survey of retracted articles from the scholarly literature. *PLoS ONE* 7(10): e44118. DOI: doi: 10.1371/journal.pone.0044118
- [11] Thed van Leeuwen and Marc Luwel. 2014. An in-depth analysis of papers retracted in the Web of Science. in *Context Counts: Pathways to Master Big and Little Data*. In *Proceedings of the science and technology indicators conference*, September 2–3, 2014, Leiden, the Netherlands, 337–344.
- [12] Martin Shepperd and Leila Yousefi. 2022. An analysis of retracted papers in Computer Science. arXiv e-prints. arXiv:2206.06706
- [13] COPE Council. Retraction Guidelines. November 2019. <https://doi.org/10.24318/cope.2019.1.4>
- [14] ICMJE Scientific Misconduct, Expressions of Concern, and Retraction. <https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/scientific-misconduct-expressions-of-concern-and-retraction.html>
- [15] Editors-in-Chief statement regarding published clinical trials conducted without IRB approval by Joachim Boldt. March 12, 2011. In. *Minerva Anesthesiologica* 2011 May;77(5):562-3 <https://www.minervamedica.it/en/journals/minerva-anesthesiologica/article.php?cod=R02Y2011N05A0562>
- [16] Retractiondatabase. <http://retractiondatabase.org/RetractionSearch.aspx#?auth%3Dboldt>; <https://retractionwatch.com/category/by-author/joachim-boldt-retractions/>
- [17] Ryan Zarychanski, Ahmed M. Abou-Setta, Alexis F. Turgeon, Brett L. Houston, Lauralyn McIntyre, John C. Marshall, and Dean A. Fergusson. 2013. Association of Hydroxyethyl Starch Administration with Mortality and Acute Kidney Injury in Critically Ill Patients Requiring Volume Resuscitation: A Systematic Review and Meta-analysis. *JAMA* 309, 7 (Feb. 2013), 678–688. DOI:10.1001/jama.2013.430
- [18] Alison Avenell, Andrew Grey, Greg Gamble, and Mark Bolland. 2019. An investigation into the impact and implications of published papers from retracted research: systematic search of affected literature. *BMJ Open* 2019;9:e031909. doi: 10.1136/bmjopen-2019-031909 [Corrections to references in this paper at <http://dx.doi.org/10.1136/bmjopen-2019-031909corr1>, October 2019]
- [19] R. Grant Steen. 2011. Retractions in the medical literature: how many patients are put at risk by flawed research? *Journal of Medical Ethics* 37, 11 (2011), 688–692.
- [20] Consolato Gianluca Nato, Leonardo Tabacco, and Federico Bilotta. 2022. Fraud and retraction in perioperative medicine publications: What we learned and what can be implemented to prevent future recurrence. *Journal of Medical Ethics* 48, 7 (2022), 479–484. Retrieved from <https://jme.bmj.com/content/48/7/479> [this article has been corrected but the errors did not affect the original conclusions.]
- [21] Peiling Wang and Jing Su. 2022. Expert recommended biomedical journal articles: Their retractions or corrections, & post-retraction citing. *Journal of Information Science*. (Feb. 2022), 8 pages. DOI: 10.1177/01655515221074329.
- [22] Alison Avenell, Mark Bolland, Greg Gamble, and Andrew Grey. 2022. Effect of alerting authors of systematic reviews and guidelines that research they cited had been retracted. The 9th Peer Review Congress: *Enhance the quality and credibility of science*, September 8–10, 2022, Chicago, IL, 24–25.
- [23] Judit Bar-Ilan and Gali Halevi. 2017. Post retraction citations in context: a case study. *Scientometrics* 113 (2017), 547–565. DOI: 10.1007/s11192-017-2242-0.
- [24] Tzu-Kun Hsiao and Jodi Schneider. 2021. Continued use of retracted papers: Temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. *Quantitative Science Studies* 2, 4, (2022), 1144–1169. DOI: https://doi.org/10.1162/qss_a_00155
- [25] Peiling Wang, Luke McCullough, and Jing Su. 2022. Continued use of retracted publications: Implications for scientific publishing & information systems. *iConference*, February 28–March 4, 2022. www.ideals.illinois.edu/items/123104
- [26] Rebecca Lea Morris. 2021. Retracted papers continue to circulate on social media, spreading misinformation. *Massive Science* (2021). <https://massivesci.com/notes/retraction-social-media-publishing-altmetric/>
- [27] Stylianos Serghiou, Rebecca M. Marton, and John P.A. Ioannidis. 2021. Media and social media attention to retracted articles according to Altmetric. *PLoS ONE* 16, 5 (May 2021). e0248625. <https://doi.org/10.1371/journal.pone.0248625>
- [28] Rosy Jan and Tazeem Zainab. 2018. The Impact story of retracted articles: Altmetric it! In *5th IEEE International Symposium on Emerging Trends and Technologies in Libraries and Information Services*, February 21–23, 2018, Noida, India, 418–422. DOI: 10.1109/ettlis.2018.8485245
- [29] Hao Peng, Daniel M. Romero, and Emöke-Ágnes Horvát. 2022. Dynamics of cross-platform attention to retracted papers. *Proceedings of the National Academy of Sciences of the United States of America* 119, 25 (June 2022). <https://doi.org/10.1073/pnas.2119086119>
- [30] Jodi Schneider, Di Ye, Alison M. Hill and Ashley S. Whitehorn. 2020. Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. *Scientometrics* 125, 2877–2913.
- [31] John Budd, Zach Coble, and Alison Abritis. 2016. An investigation of retracted articles in the biomedical literature. *Proceedings of the ASIST* 53 (2016), 1–9. DOI: 10.1002/pra2.2016.14505301055.
- [32] Jaime A. Teixeira da Silva and Helmar Bornemann-Cimenti. 2016. Why do some retracted papers continue to be cited?. *Scientometrics* 110 (2017), 365–370. <https://doi.org/10.1007/s11192-016-2178-9>
- [33] Enago Academy. 2022. Database of Retracted Papers Launched for Researchers. (Apr. 2022). Retrieved October 1, 2022 from <https://www.enago.com/academy/database-of-retracted-papers-launched-for-researchers/>
- [34] Mark P. Pfeifer and Gwendolyn L. Snodgrass. 1990. The Continued Use of Retracted, Invalid Scientific Literature. *JAMA* 263, 10 (1990), 1420–1423.
- [35] Jaime A. Teixeira da Silva. 2022. A Synthesis of the Formats for Correcting Erroneous and Fraudulent Academic Literature, and Associated Challenges. *Journal for General Philosophy of Science*. <https://doi.org/10.1007/s10838-022-09607-4>
- [36] Jodi Schneider, Nathan D. Woods, Randi Proescholdt, Yuanxi Fu, and The RISRS Team. Recommendations from the Reducing the Inadvertent Spread of Retracted Science: Shaping a Research and Implementation Agenda Project. July 2021. MetaArXiv Preprints. <http://doi.org/10.31222/osf.io/ms579>
- [37] Quan-Hoang Vuong. 2020. Reform retractions to make them more transparent *Nature* 582, 7811 (2020), 149. DOI: <https://doi.org/10.1038/d41586-020-01694-x>

- [38] Jaime A. Teixeira da Silva and Quan-Hoang Vuong. 2022. Fortification of retraction notices to improve their transparency and usefulness. *Learned Publishing* 35, 292-299. <https://doi-org.utk.idm.oclc.org/10.1002/leap.1409>
- [39] Daniele Fanelli. 2016. Set up a 'self-retraction' system for honest errors. *Nature* 531, 415 (2016). <https://doi.org/10.1038/531415a>
- [40] Quan-Hoang Vuong. 2020. The limitations of retraction notices and the heroic acts of authors who correct the scholarly record: An analysis of retractions of papers published from 1975 to 2019. *Learned Publishing* 33, 119-130. <https://doi-org.utk.idm.oclc.org/10.1002/leap.1282>
- [41] Elizabeth M. Suelzer, Jennifer Deal, Karen Hanus, Barbara E. Rugeri, and Elizabeth Witkowski. 2021. Challenges in Identifying the Retracted Status of an article. *JAMA Network Open*. 2021;4(6):e2115648. doi:10.1001/jamanetworkopen.2021.15648
- [42] Caitlin Bakker and Amy Riegelman. 2018. Retracted Publications in Mental Health Literature: Discovery across Bibliographic Platforms. *Journal of Librarianship and Scholarly Communication* 6, 1 (2018). DOI: <https://doi.org/10.7710/2162-3309.2199>
- [43] Jaime A. Teixeira da Silva and Serhii Nazarovets. 2022. Publication History: A Double-DOI-Based Method for Storing and/or Monitoring Information about Published and Corrected Academic Literature. *Journal of Scholarly Publishing* 53, 2 (Jan. 2022), 85-108.
- [44] Richard van Noorden. 2011. Science publishing: The trouble with retractions: A surge in withdrawn papers is highlighting weaknesses in the system for handling them. *Nature* 478, 7367 (2011), 26–28. <https://doi.org/10.1038/478026a>
- [45] Elizabeth C. Moylan and Maria K. Kowalczyk. Why articles are retracted: a retrospective cross-sectional study of retraction notices at BioMed Central. *BMJ Open* 2016;6:e012047. DOI: 10.1136/bmjopen-2016-012047
- [46] Robin Haunschild and Lutz Bornmann. 2021. Can tweets be used to detect problems early with scientific papers? A case study of three retracted COVID-19/SARS-CoV-2 papers. *Scientometrics* 126 (2021), 5181–5199. <https://doi.org/10.1007/s11192-021-03962-7>
- [47] Everett M. Rogers. 2003. *Diffusion of Innovations* (5th ed.). Free Press, New York, NY.

APPENDICES

A ANALYSIS OF A CASE OF RETRACTING AN ARTICLE BY A JOURNAL

In this example, we observe this highly-cited article (received 1097 citations in Scopus and 995 citations in WoS) to illustrate the process of retraction:

1. The original article was published in *The Lancet* on November 14, 2011. DOI: 10.1016/S0140-6736(11)61590-0
2. *Expression of Concern: the SCIPIO trial* was posted by Lancet Editors on April 12, 2014. DOI: 10.1016/S0140-6736(14)60608-5
3. *Retraction* published by Lancet Editors on March 16, 2019, DOI: 10.1016/S0140-6736(19)30542-2. It states:

“The results of these investigations persuade us that the laboratory work undertaken by Piero Anversa and colleagues at Harvard cannot be held to be reliable. Specifically, there are issues with the data presented in figures 2 and 3 and supplemental figures 2 and 3. SCIPIO was a collaboration between Anversa’s laboratory in Boston, MA, USA, and Roberto Bolli’s team in Louisville, KY, USA. Anversa’s laboratory isolated, expanded, and characterised the c-kit positive cells, which were then shipped to Louisville, where they were administered to patients and all the clinical work was done. The Louisville team was not involved with the manufacturing and characterisation of c-kit positive cells. Although we do not have any reservations about the clinical work in Louisville that used the preparations from Anversa’s laboratory in good faith, the lack of reliability regarding the laboratory work at Harvard means that we are now retracting this paper.”

How are the three documents linked? Document 1 has two links to documents 2 and 3 under the section Linked Articles. However, from document 2, there is only a link to document 1 (Should it link to Document 3 the decision?). Document 3 only links to document 1, not document 2 (the history was not preserved). The practice is to link backward but not forward. From the user’s perspective, a link from document 2 to document 3 is also necessary to follow up on document 2’s promise: “As soon as *The Lancet* receives further information, we will inform readers accordingly.”

How are the article’s HTML and PDF displayed? From the original article Webpage, both HTML full text and PDF articles were clearly stamped as “RETRACTED” but the title does not have “Retracted.” Using the Export Citations, the entry is in Figure 4

How do databases retrieve the three related documents? In WoS and Scopus, we did a simple search of the first three authors. We only retrieved the “Retracted:” article. The 964 citing publications were distributed across three periods:

1. Between publication and Expression of Concern
2. Between Expression of Concern and Retraction
3. Post-retraction

When we searched the title of the original article “Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO)”, we retrieved both the “Retracted:” article and the “RETRACTION:” notice, not the Expression of concern. Why was the record for “Expression of concern” not in the results from author or title queries? Searches using DOI located the records and solved the puzzle too:

1. The Expression of Concern is titled “Expression of concern: the SCIPIO trial”
2. The original article is titled “Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO): initial results of a randomised phase 1 trial”
3. The document type for the record is “Editorial Material” (WoS now has “Expression of concern” as an index term for Document Type but not at the time this record was indexed)

We further searched the free databases Crossref and controversial free articles site Sci-Hub. Crossref has the record linked to the publisher’s pdf with the watermark “Retracted” but the bibliographic record does not have “Retracted” in the title, nor any links to the Expression of Concern or Retraction Notice. As of December 30, 2022, Sci-Hub has direct access to the PDF article that does not have a “Retracted” watermark at [https://sci-hub.se/10.1016/S0140-6736\(11\)61590-0](https://sci-hub.se/10.1016/S0140-6736(11)61590-0)

The case can be summarized:

1. Retracting a published article can be a long process.
2. A set of documents are related; they contain different information of importance to science and users.
3. When an article is being investigated under “Expression of Concern,” its final status is uncertain.

IRs face challenges to integrate the set of related documents in databases and need search features for these documents.

To minimize and prevent post-retraction use of invalid science, IIR should develop strategies to inform users about the retractions and corrections of the articles they have accessed and collected

Export Citations

Export citations in RIS format, which can be used with EndNote, ProCite, RefWorks, and most other reference management software.

Options

Citation for the content listed below

Citation + Abstract

Content

1. **Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO): initial results of a randomised phase 1 trial**
Bolli, Roberto et al.
The Lancet, Volume 378, Issue 9806, 1847 - 1857

Figure 4: Lancet save record

B CORRECTIONS OF MULTIPLE ARTICLES

In Figure 5, the 4th correction is for the example in 2.2.5. This page is a subscription or pay-for-access. The last sentence states “Additional information is available at [\[full/science.1188888/DC2\]\(http://www.sciencemag.org/cgi/content/full/science.1188888/DC2\).” However, the URL will redirect to the article page at \[www.science.org/doi/10.1126/science.1188888\]\(http://www.science.org/doi/10.1126/science.1188888\). Readers need to download the File \(wyatt_clarification.pdf\) to view additional information mentioned in the Correction Notice. Although this downloading is free, it is not easy to find.](http://www.sciencemag.org/cgi/content/</p></div><div data-bbox=)

CORRECTIONS AND CLARIFICATIONS

Cover Caption: (17 June, p. 1349). The caption identified the location of the image as Rikuzentakada, Miyagi Prefecture. Rikuzentakada is in the Iwate Prefecture.

Perspectives: "Innate immunity in plants goes to the PUB" by L. A. J. O'Neill (17 June, p. 1386). In the figure and legend, the RD kinases BAK1 and BIK1 in *Arabidopsis*, and the equivalent kinases IRAK-1 and IRAK-2 in humans, were mislabeled as non-RD kinases. Pattern recognition receptors in plants such as FLS-2 associate with or carry kinases of the non-RD subclass to transduce early signaling events in innate immunity. They typically partner with RD kinases. How the RD kinases interact with non-RD kinases remains unknown.

Research Articles: "EPOXI at Comet Hartley 2" by M. F. A'Hearn *et al.* (17 June, p. 1396). An error in the author list and affiliations was introduced in proofs. The author list should have included two authors named Michael S. Kelley. One is affiliated with the Department of Astronomy, University of Maryland, College Park, MD 20742–2421, USA. The other is affiliated with the Planetary Science Division, NASA Headquarters, Mail Suite 3V71, 300 E Street SW, Washington, DC 20546, USA. The error has been corrected in the online PDF and HTML versions.

Research Article: "*Staphylococcus aureus* nonribosomal peptide secondary metabolites regulate virulence" by M. A. Wyatt *et al.* (16 July 2010, p. 294). During the construction of the *ausA* (responsible for making dipeptide secondary metabolites) deletion strain, an inadvertent secondary site mutation in the *sae* two-component regulator gene *saeS* occurred, as determined by subsequent genome sequencing. This incorrectly led to an association of the described aureusimine cyclic dipeptides with gene transcription affected by the *sae* secondary site mutation. Additional information is available at www.sciencemag.org/cgi/content/full/science.1188888/DC2.

Reports: "Skyrmion lattice in a chiral magnet" by S. Mühlbauer *et al.* (13 February 2009, p. 915) and "Spin transfer torques in MnSi at ultralow current densities" by F. Jonietz *et al.* (17 December 2010, p. 1648). The authors clarify that the temperatures reported have an absolute uncertainty of less than 5% because the temperature values were determined with batch-calibrated Pt1000 thermometers (accuracy ± 1 K). In those experiments where the thermometer was not directly attached to the sample, a small additional temperature gradient between sample and thermometer was present. This explains small discrepancies in the location of the reported data points compared with the phase boundaries reported elsewhere. All relative temperatures are perfectly consistent with other papers, and therefore neither the identification of the A phase nor any other conclusions of the publications are affected.

9 SEPTEMBER 2011 VOL 333 SCIENCE www.sciencemag.org
 Published by AAAS

Science

Corrections and Clarifications

Science, 333 (6048), • DOI: 10.1126/science.333.6048.1381-b

Figure 5: A section for corrections of 5 documents