

A Multidimensional Investigation of the Effects of Publication Retraction on Scholarly Impact

Xin Shuai 

*Research & Development Group, Thomson Reuters, 610 Opperman Dr., St. Paul, MN 55123.
E-mail: xin.shuai@thomsonreuters.com*

Jason Rollins 

*Clarivate Analytics, 50 California St., San Francisco, CA 94111.
E-mail: jason.rollins@thomsonreuters.com*

Isabelle Moulinier

*Research & Development Group, Thomson Reuters, 610 Opperman Dr., St. Paul, MN 55123.
E-mail: isabelle.moulinier@thomsonreuters.com*

Tonya Custis

*Research & Development Group, Thomson Reuters, 610 Opperman Dr., St. Paul, MN 55123.
E-mail: tonya.custis@thomsonreuters.com*

Mathilda Edmunds 

*Clarivate Analytics, 1500 Spring Garden St., Philadelphia, PA, 19130.
E-mail: mathilda.edmunds@thomsonreuters.com*

Frank Schilder

*Research & Development Group, Thomson Reuters, 610 Opperman Dr., St. Paul, MN 55123.
E-mail: frank.schilder@thomsonreuters.com*

During the past few decades, the rate of publication retractions has increased dramatically in academia. In this study, we investigate retractions from a quantitative perspective, aiming to answer two fundamental questions. One, how do retractions influence the scholarly impact of retracted papers, authors, and institutions? Two, does this influence propagate to the wider academic community through scholarly associations? Specifically, we analyzed a set of retracted articles indexed in Thomson Reuters Web of Science (WoS), and ran multiple experiments to compare changes in scholarly impact against a control set of nonretracted articles, authors, and institutions. We further applied the

Granger Causality test to investigate whether different scientific topics are dynamically affected by retracted papers occurring within those topics. Our results show two key findings: first, the scholarly impact of retracted papers and authors significantly decreases after retraction, and the most severe impact decrease correlates with retractions based on proven, purposeful scientific misconduct; second, this retraction penalty does not seem to spread through the broader scholarly social graph, but instead has a limited and localized effect. Our findings may provide useful insights for scholars or science committees to evaluate the scholarly value of papers, authors, or institutions related to retractions.

Received February 25, 2016; revised September 12, 2016; accepted September 27, 2016

© 2017 The Authors. Journal of the Association for Information Science and Technology published by Wiley Periodicals, Inc. on behalf of ASIS&T • Published online 26 April 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23826

Introduction

Sir Isaac Newton said, “If I have seen further, it is by standing on the shoulders of giants.” Scientific progress and communication, mostly presented in the form of scientific publications, have laid a solid foundation for the overall

development and advance of science for centuries. Accumulative scientific advances are predicated on the assumption that scholars are honest and serious about the accuracy and integrity of their published work. Unfortunately, this is not always true. The scientific community bears the responsibility of self-examination and self-correction to ensure the integrity and authority of scientific literature.

During the past few decades, the rate of paper retractions, because of errors or purposeful misconduct, has increased dramatically in nearly all academic fields (Noorden, 2011; Steen, Casadevall, & Fang, 2013). A retraction of a published article indicates that the ideas, methodology, or results presented in the original article are shown to be scientifically invalid, and therefore can no longer serve as the proverbial “shoulders of giants.” The most common reasons for retraction are scientific misconduct (i.e., falsification, plagiarism) or unintended errors. Retractions are typically initiated by journal editors or by the article’s authors themselves.

Recently, a growing list of retracted papers has drawn the attention and scrutiny of both the academic and the popular media. In one widely reported example, a study about evolving attitudes toward gay marriage published in *Science* (LaCour & Green, 2014) was retracted because one of the authors was not able to provide the raw data. *The Lancet* retracted a study by Wakefield et al. (1998) that suggested that combined vaccines of measles, mumps, and rubella lead to autism in children. Despite the retraction of the study, many parents continue to believe it, which has resulted in a decline of vaccines for children in Britain and the United States. In another prominent case, two papers about human stem cells published in *Science* (Hwang, Roh, et al., 2005), were retracted because the authors fabricated the data.

Although the systematic and exhaustive study of the publication retraction is still in a nascent stage, several previous studies explore this area from different aspects and inform our work. Fang, Steen, and Casadevall (2012) categorized more than 2,000 biomedical and life science research articles based on retraction reasons and found that more than 60% of retractions are attributed to misconduct. Lu, Jin, Uzzi, and Jones (2013) conducted a controlled experiment showing that the citation counts of retracted articles decrease more rapidly after retraction than those for nonretracted control articles, and prior articles by authors of retracted papers, published before retraction, are also negatively affected. Chen, Hu, Milbank, and Schultz (2013) discussed the use of a visualization tool to study the relation of retracted articles to other scientific literature.

As an extension of the prior studies, the main focus of our research is to address two problems: One, how do retractions affect the scholarly impact of papers, authors, and institutions? Two, does such an effect propagate to the wider scientific or academic community through scholarly association? The main contributions of this paper include the following:

- We designed multiple statistical comparative experiments within a unified framework on a 30-years span of retracted publications indexed by Web of Science (WoS) to study the

effect of retractions on the scholarly impact of articles, scholars, and institutions, which are either directly involved in or indirectly related to multiple categories of retraction incidences. We also compared our experimental results with similar previous studies.

- We investigated temporal causal correlation between paper retraction rate and publishing popularity within certain scientific topics. We test whether retraction occurrences in one research direction will affect its future development.

Related Work

The phenomenon of publication retraction in academia has recently increased. Although the scientific investigation of retraction is still at an early stage, several studies attempt to explore the characteristics of retracted papers and their effect on scholarly impact.

Descriptive Analysis and Characterization of Retraction

Although the increasing number of retraction cases have received attention from several scholars, the scientific community has yet to pay full attention to it (Elizabeth, 2015; Noorden, 2011), and retracted papers are still being cited (Pfeifer & Snodgrass, 1990). With regard to the reason why the number of retractions has increased, Steen et al. (2013) examined the interval between publication and retraction for 2,047 retracted articles indexed in PubMed. They concluded that the recent increase in retraction reflects the changes in both institutional policy and author behavior. The phenomenon of retraction itself appears to be a relatively recent phenomenon for most journals; authors responsible for multiple retractions have a considerable impact on most recent retractions; especially, serial misconduct has been identified more frequently in recent years. Fanelli (2013) further pointed out that the increase of retraction is actually a good sign for science because their analysis reveals that retractions have grown not because of rising misconduct, but because researchers and journal editors are getting better at identifying fraudulent publications.

The categorization of retraction reasons has been studied by many researchers. Fang et al. (2012) categorized retraction reasons for more than 2,000 biomedical and life science research articles by consulting multiple resources in addition to retraction notices and found that more than 60% of retractions were attributed to misconduct. Grieneisen and Zhang (2012) did a similar analysis and emphasized that several prolific authors accounted for more than half of all retractions as a result of alleged research misconduct, and strongly influenced all retraction characteristics. Madlock-Brown and Eichmann (2015) proposed a more complete categorization schema and found out that most citations of retracted papers after retraction are from the authors themselves.

In addition to categorization, Chen et al. (2013) discussed the use of a visualization tool to study relationships between retracted articles and other scientific literature. In addition, there are several studies that attempt to find factors that

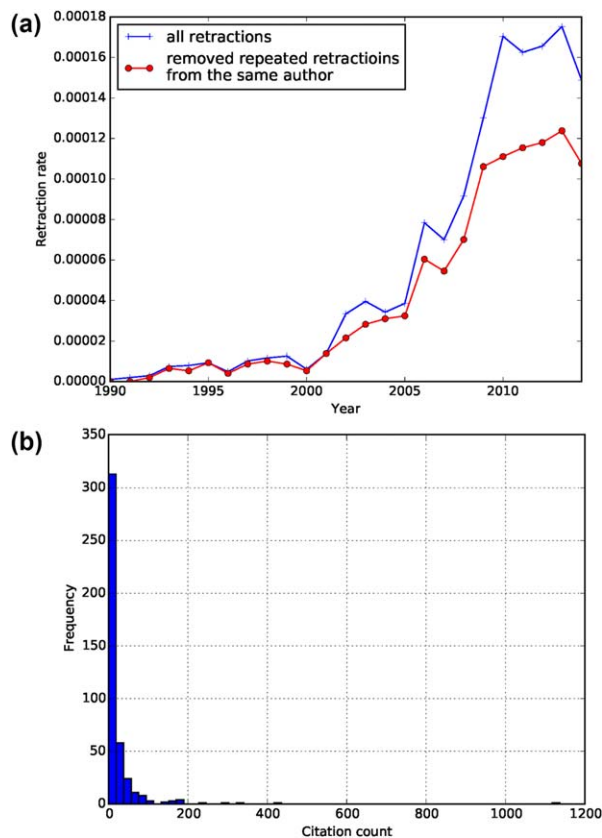


FIG. 1. Several statistics about retracted papers. Specifically, (a) shows the dynamic change of retraction rate and (b) shows the distribution of citation counts of retracted papers. [Color figure can be viewed at wileyonlinelibrary.com]

correlate with retractions. Furman, Jensen, and Murray (2012) suggest that attention is a key predictor of retraction, and that retractions arise most frequently among highly cited articles. Fanelli, Costas, and Larivière (2015) point out that the key factors leading to scientific misconduct are lack of research integrity policies, money-driven publishing behavior, and career pressure for junior researchers.

Quantitative and Statistical Analysis of Impact As a Result of Retraction

Many studies describe and quantify the aftermath of paper retractions. Some researchers focus on specific retraction cases. Bornemann-Cimenti, Szilagyi, and Sandner-Kiesling (2016) examine the citation change of scholar Scott Reuben's articles within 5 years of their retraction, and found that most of this retracted work is still being cited without mention of retraction by other scholars. Michalek, Hutson, Wicher, and Trump (2010) defined a mathematical equation to measure the potential economic cost of scientific misconduct by a federal grant applicant.

Other researchers extended the investigation from specific cases to a set of retracted articles or authors. Lu et al. (2013) compared the citation counts after retraction between retracted articles and nonretracted control ones through

statistical tests, and found that the former decrease more rapidly than the latter. They also showed that the citation counts of prior-retraction publications by authors of retracted papers, are significantly less than controlled ones. Azoulay, Furman, Krieger, and Murray (2014) applied a similar methodology to a set of articles that were nonretracted but topically related to retracted articles and found that the penalty also applied to related articles, but the penalty was not as substantial as for the retracted articles. Azoulay, Bonatti, and Krieger (2015) applied a similar controlled experiment methodology and showed evidence for the negative effects of retractions on papers and their authors' citation count. They also showed that retractions as a result of fraud lead to a more severe negative effect than honest errors. Mongeon and Larivière (2015) further confirmed the negative impacts to coauthors involved in a retracted article.

Our work aligns with and extends these lines of research by examining the effects of retraction on the scholarly impact of articles, scholars, and institutions from various aspects within a unified statistical experimental framework. We further investigate whether such an effect propagates to the wider scientific or academic community through scholarly associations.

Data Description and Overview

Our dataset comprises records from Thomson Reuters WoS based on articles published from 1980 to 2014 across a wide range of research disciplines. For each article, metadata including Title, Authors, Published Date, Journal, ESI category,¹ Institutions, and Cited References were used in our analysis. To identify retracted publications, we searched for titles containing "Retracted article" (e.g., *Arabidopsis Downy Mildew Resistance Gene RPP27 Encodes a Receptor-like Protein Similar to CLAVATA2 and Tomato Cf-9* [Retracted Article. See vol. 143, p. 1079, 2007] Tör et al. [2004]).

In total, we extracted 2,659 retracted articles. The change of annual retraction rate, which we define as the ratio of the number of retracted articles to the total number of published articles in a certain year, is shown in Figure 1a. Because several studies (Grieneisen & Zhang, 2012; Steen et al., 2013) mentioned the effect of prolific authors on retraction statistics, we removed repeated retractions from the same authors and recalculated the annual retraction rate. From this, we see that the annual retraction rate has unmistakably increased over the past 25 years, regardless of whether repeated retractions are removed. However, the most recent retraction rate has been clearly inflated because of the detection of multiple retractions from the same authors. Table 1 shows the top five, as well as bottom five scientific ESI categories ranked by retraction ratio (i.e., the number of papers retracted over the total number of papers published in that category).

¹ESI category is a journal subject categorization system coined by Thomson Reuters; see http://ipsience-help.thomsonreuters.com/incitesLive/7622-TRS/version/1/part/5/data/ESI_Journal_Category_Map_2012.xlsx?branch=incites_115&language=en_US

TABLE 1. Top five and bottom five frequently retracted research fields defined in ESI categories.

Rank	ESI category	Retraction rate ($\times 10^{-4}$)
1	Molecular biology & genetics	2.1
2	Immunology	1.66
3	Microbiology	1.55
4	Biology & biochemistry	1.14
5	Pharmacology & toxicology	1.08
18	Plant & animal science	0.36
19	Social sciences, general	0.21
20	Geosciences	0.17
21	Computer science	0.17
22	Space science	0.06

Notably, high retraction rates mostly occur in medical or biological-related research fields. Figure 1b shows the distribution of the citation counts of retracted articles and the distribution is very skewed. We also found that the median citation count of retracted articles is eight, whereas the median count for all articles is only one, indicating that a retracted article is generally cited more than an average article.

As to the reason for the increased speed of retractions, several studies have proposed their interpretations. Resnik, Wager, and Kissling (2015) conducted a survey analysis which reveals that three times as many journals in their sample now have a retraction policy compared to an earlier study in a similar group of journals (Atlas, 2004); this seems to have contributed, at least in part, to the increasing findings of retraction cases in recent years. Steen et al. (2013) analyzed several factors that may be correlated with the increase of retractions, and attributed the trend to changes both in institutional policy and author behavior: the phenomenon of retraction itself appears to be a relatively recent phenomenon for most journals; authors responsible for multiple retractions have a considerable impact on most recent retractions; especially serial misconduct has been identified more frequently in recent years. Figure 1a also confirms their findings. Fanelli (2013) demonstrated that the growth of retractions is mainly a result of the increasing number of journals issuing retractions, but not a sign of prevalent behaviors of misconduct. Inspired by their work, we hypothesize that the increase of journals retracting articles may also contribute to the increase of total retraction cases, and leave the verification of the hypothesis in our future work.

Concepts and Definitions

This section is organized as follows: First, we design a coding schema to categorize the reasons for retraction, which enables us to investigate how different types of retraction lead to different degrees of effect on scholarly impact. Second, we define retracted papers, authors, and institutions, as well as related papers and authors, in terms of citation relations. Finally, we describe the concepts of research topics, topical popularity, and retraction rate to examine

possible temporal correlations between retractions and topic popularity.

Coding Schema for Retraction Categorization

Scholarly publications are retracted for several reasons. Typically, either the authors of the paper or the editor of the publishing journal can request a retraction. Understanding the reason for retraction is useful in scholarly impact analysis; we hypothesize that different types of retraction have different degrees of effect on postretraction impact. After a preliminary investigation of a sample set of retracted papers, we based our retraction classifications schema on Grieneisen and Zhang (2012) and Fang et al. (2012), but streamlined it and factored in our understanding of the generally accepted view of the primary reasons for retraction as follows:

- Scientific misconduct: the intentional violation of standard codes of scholarly conduct and ethical behavior in scholarly publication of scientific research. It can be further divided into three subcategories:
 - Plagiarism: copying the ideas, contents (including text, figures, tables, data), and results of others without explicit citation; duplicate publication of the same materials in different journals.
 - Falsification or Fabrication: making up results or data, manipulating research materials, process, or data so that the research is not accurately represented.
 - Violation of rules: submitting articles without approval of all coauthors; conducting human experiments without Institutional Review Board (IRB) approval; or other types of violations not directly related to the content of articles.
- Errors: unintentional errors made by authors during the process of data collection, processing, or analysis causing the final results to be invalid. Errors can also occur during the process of article publishing.
- Others: any other reasons that are not the fault of the authors. For instance, the publishers may unintentionally publish two versions of the same article and need to retract one version.
- Not found: the retraction notice was not found or the reason for retraction was not explicitly given.

In addition, we categorized the following three types of retraction requests: editor's request, author's request, or not found.

Four raters, including three authors of this study, annotated 1,666 of 2,659 retracted articles, tagging who requested the retraction and the reason for retraction. To quantify the interrater agreement, a set of 100 articles are randomly selected and assigned for annotation to all four raters. The calculated Fleiss's kappa is .73, indicating that the degree of interrater agreement is high and the annotation results are trustworthy. Figure 2 shows the distribution of reasons for retraction, based on articles or authors. More than 25% of retraction cases are because of plagiarism (including duplicate publications). The second-most prevalent reason for retractions (around 24%) is unintentional author errors in the process of experimentation or data analysis. The third-most prevalent reason for retractions (around 23%) is falsification

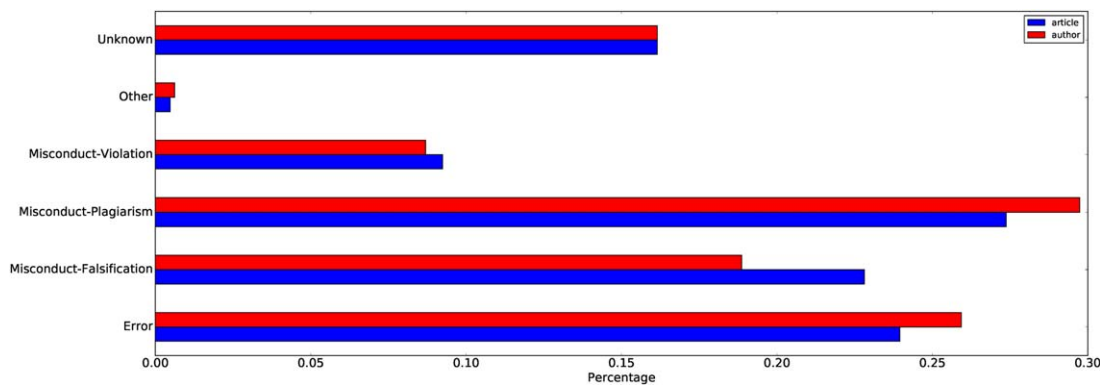


FIG. 2. The distribution of retraction reasons for articles and authors. The most frequent reason for retraction is plagiarism, followed by error and falsification. [Color figure can be viewed at wileyonlinelibrary.com]

and fabrication. Both plagiarism and falsification are considered egregious forms of scientific misconduct. A clear difference in distribution of reasons between articles and authors is that the percentage of falsification and fabrication of the former is larger than the latter. It indicates that, once an author was found to fabricate and falsify data in one article, a series of related articles from the same author were affected and retracted as well.

The analysis of retraction reasons reveals that the majority of retractions are a result of scientific misconduct (i.e., around 60%), which corresponds to the major finding by Fang et al. (2012). However, the percentage of falsification/fabrication type of misconduct in our dataset (i.e., 22.8%), is only half of that in their dataset (i.e., 43.4%). We attribute this difference to two possible reasons. One, all of the articles they analyzed are indexed by PubMed, whereas around 15% to 20% articles we studied are not indexed by PubMed because we used the WoS dataset. Second, they consulted several secondary sources to double-check the retraction reason in addition to retraction notices. Although we made similar efforts, in the vast majority of cases we were simply not able to find any additional reliable resources beyond the original retraction notice from the publisher.

Retracted Entities and Related Entities

First, we define the main concepts as follows:

- **Academic entity:** Paper, author, or institution in the WoS corpus. Each paper is assigned a unique ID, and authors and institutions are each represented by a unique name string after disambiguation and normalization.
- **Scholarly impact:** Total number of citation counts. Scholarly papers are connected through citations, and the citation count is one of the most recognized indicators of scholarly impact. Similarly, the scholarly impact of authors and institutions can be represented by the sum of citation counts of their corresponding papers.
- **Impact curve:** The temporal curve of yearly citation counts. If we use $C(y)$ to denote the scholarly impact of any entity in year y , then the *impact curve* can be formulated as a temporal curve within specific period: $C(y), y_0 \leq y \leq y_n$. We set

y_n to 2014, and y_0 has different definitions based on the type of entity. For each paper, y_0 is the publication year; for each author or institution, y_0 is the first year when the author or institution is cited, i.e., the year that the scholarly impact begins to build up.

- **Retraction year:** The first year when an academic entity was involved in a retraction. For different entities, the specifications of retraction year are different, and the details will be shown later.
- **Retracted entity:** Academic entity that was involved in a retraction case. Specifically, once a paper was announced as retracted, all authors of this paper, as well as the institutions those authors were affiliated with, are considered retracted authors and retracted institutions.

Scholarly impact is not a static measure, but dynamically changing all the time as new publications emerge. The main question under investigation in this paper is how formal retractions affect the scholarly impact of involved and related papers, authors, and institutions.

Scholar and Institution Name Disambiguation

Each article extracted from the Thomson Reuters WoS collection has a unique ID field. In contrast, authors and institutions are represented only by name strings, which often need to be disambiguated.

We use two proprietary Thomson Reuters tools for disambiguation. The first tool (Griffith, 2015) uses a semisupervised machine learning algorithm that clusters all author names in WoS into different groups, with each group corresponding to one unique scholar; this reaches 95% precision and 84% recall. For institution name disambiguation, we used the Web Application for Address Normalization (Intellectual Property & Science, 2013), which normalizes to the root-organization level (e.g., University), but not to the sub-organization level (e.g., Department or School).

One Retraction Example

Two papers about cloning and human stem cells were published in *Science* (Hwang, Roh, et al., 2005; Hwang, Ryu, Park, Park, & Lee, 2004) and retracted in 2006,

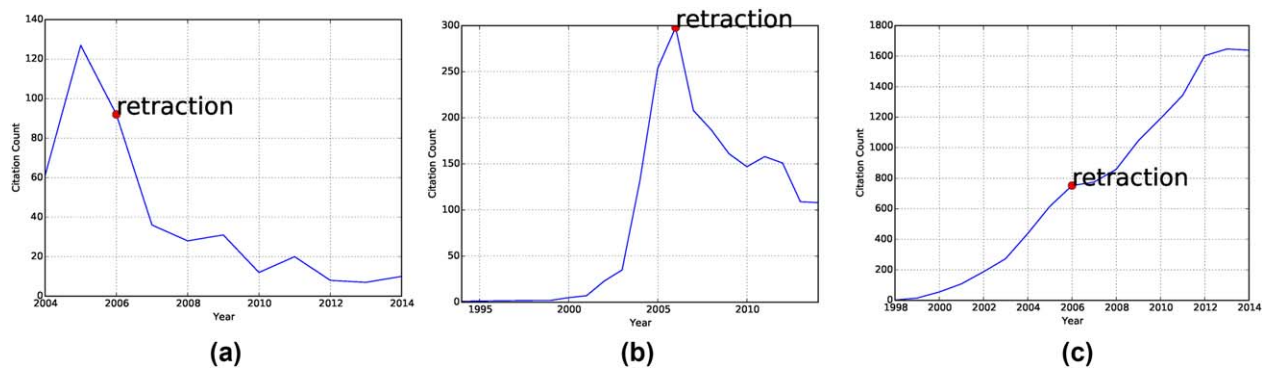


FIG. 3. The impact curves of (a) the 2004 stem cells paper and (b) Hwang, Woo Suk, and (c) Seoul National University College of Veterinary Medicine. In (a) and (b) we can spot the decreasing trend after the retraction point. [Color figure can be viewed at wileyonlinelibrary.com]

because much of the data were found to be fabricated. The name of both papers' first author is Hwang, Woo Suk, who was a professor at Seoul National University College of Veterinary Medicine. Figure 3 shows the impact curves for the retracted paper published in 2004, the author, and the institution. We can clearly spot the decrease of scholarly impact after retraction for the retracted paper and the author (see Figure 3a,b). However, such a decrease is not shown for the institution (see Figure 3c). We conduct statistical comparative experiments to test the significance of these decreases in citation count below.

Related Entities to Retraction

In addition to the explicitly retracted entities, effects on other academic entities that are related to the retracted entities are also within the scope of our investigation. For retracted papers, other papers that cite the retracted papers (i.e., citing relation) or share the same references as the retracted papers (i.e., coreference relation) are considered to be related papers. A related author (i.e., coauthorship relation) is defined as an author who once coauthored with a retracted author, but on a nonretracted paper. We do not consider related institutions here because it is not easy to define and measure the relatedness between institutions. By exploring changes in scholarly impact of related, but nonretracted, papers and authors, we attempt to address whether the retraction effect spreads through the scientific community through scholarly association. Table 2 summarizes the six types of retracted and related entities that will be examined.

TABLE 2. Six types of entities that will be investigated on their scholarly impact change after retraction.

Entity	Variable/symbol
Retracted papers	P
Retracted authors	A
Retracted institutions	I
Papers citing retracted papers	P_{citing}
Papers sharing references with retracted papers	P_{coref}
Authors coauthoring with retracted authors in nonretracted papers	A_{coaut}

Research Topics and Retraction

The academic entities illustrated above in Table 2 can also be grouped by scientific topics. Here, the scientific topic refers to a specific research direction within a more general scientific discipline (defined by ESI category). For instance, *Stem Cell* is a hot topic within *Molecular Biology*, and *Big Data* is an emerging topic in *Computer Science*. Some key concepts are listed as follows:

- **Scientific topic:** Given a paper, we assign scientific topics based on the title of that paper; we use a short text annotation tool *Tagme*² to detect Wikipedia concepts from all titles of retracted papers. One Wikipedia concept corresponds to one Wikipedia page, and the title of the page is the name of the detected topic. We assume that all extracted topics constitute a topic set K .
- **Yearly publication count:** Given a year y , the total number of papers found in our WoS collection in this year is denoted as $Pub(y)$.
- **Yearly topical popularity:** Given a year y and a topic $k \in K$, the total number of papers that belong to topic k is denoted as $Pub^k(y)$. Then the topical popularity of topic k , can be measured as $Pop^k(y) = Pub^k(y) / Pub(y)$.
- **Yearly topical retraction rate:** Given a year y , the total number of retracted papers that belong to topic k is denoted as $Pub^{rk}(y)$. Then we can define a yearly topical retraction rate of topic k as $Ret^k(y) = Pub^{rk}(y) / Pub(y)$.

From the titles of the 2,659 retracted articles we studied, we extracted more than 4,000 Wikipedia concepts (i.e., scientific topics). For each extracted topic, we assign its ESI category as the most frequently appearing ESI category from all of its associated papers. The top 10 concepts in terms of frequency are listed in Table 3. For each topical keyword k , we construct two time series, the yearly retraction rate $Ret^k(y)$ and the yearly topical popularity $Pop^k(y)$. Figure 4 shows $Ret^k(y)$ and $Pop^k(y)$ for the top two frequently occurring topics belonging to different ESI categories: *apoptosis* and *regulation of gene expression*. Although there's no obvious temporal correlation or pattern between

²<http://tagme.di.unipi.it/>

TABLE 3. Top 10 most frequent scientific topics from retracted papers.

Topic keyword	ESI category	Frequency
Regulation of gene expression	Biology & biochemistry	107
Gene expression	Biology & biochemistry	92
Protein	Biology & biochemistry	91
Apoptosis	Clinical medicine	88
Cell (biology)	Clinical medicine	77
Gene	Molecular biology & genetics	68
Enzyme inhibitor	Clinical medicine	52
Cancer	Clinical medicine	48
Tumor	Clinical medicine	41
Inflammation	Clinical medicine	38

$Pop^k(y)$ and $Ret^k(y)$ shown in Figure 4, we examine their possible temporal correlation through statistical significance testing in the next section.

Experimental Design and Results

In this section, we conduct a series of statistical experiments to examine and quantify the effects of retraction from various aspects. First, we design comparative experiments to test the effects of retraction on three types of academic entities, that is, papers, authors, and

institutions. Second, we compared our experimental results with previous similar studies. Third, we study the Granger-causality correlation between the occurrences of retraction within a scientific topic, and the future popularity of that topic.

Statistical Comparative Experiments for Retracted or Related Entities

Statistical comparative experiments are used here to test the effect of retraction on the entities listed in Table 2: $P, A, I, P_{citing}, P_{coref}, A_{coaut}$. Although the implementation details are different for each of the six entities in Table 2, the basic steps used to design and conduct these comparative experiments are:

1. Select treatment entity type from Table 2.
2. Select treatment entities from pool of retracted entities.
3. Specify the retraction year.
4. Select control entities based on impact curve before the retraction year.
5. Run statistical test to determine if the difference between treatment and control entities is statistically significant.

Selecting Treatment Entities and Specifying Retraction Year

Table 4 shows the criteria for selecting the treatment entities group E^t and the retraction year for each type of entity. Although the methods of selecting treatment groups differ, we pick entities that are most likely to be affected by the publication retraction intuitively. For instance, the first author of a retracted paper is presumably more affected than the other authors. The paper sharing the most references with a retracted paper is an indicator of the highest topical similarity, thus more likely to be affected as well. The most frequent coauthorship relation with a retracted author is a sign of the closest scholarly relationship with that author, thus more likely to be questioned by academic peers.

We illustrate how treatment entities are selected using a typical example. The retracted paper (Hwang et al., 2004) itself is assigned to P^t . The first author Hwang, Woo Suk and his corresponding institution Seoul National University College of Veterinary Medicine are assigned to A^t and I^t , respectively. From all articles that cited the retracted paper, the first one published (Eridani, Sgaramea, & Cova, 2004) is assigned to P^t_{citing} . From all articles that share references with the retracted paper, we select the one with the largest Jaccard coefficient between both reference sets (Tabar & Studer, 2002) and assign it to P^t_{coref} . Finally, we find a researcher named Kang, Sung Keun who coauthored most frequently with Hwang, Woo Suk and assign him to A^t_{coaut} . The difference between the two scholars is that Kang, Sung Keun did not publish any retracted works.

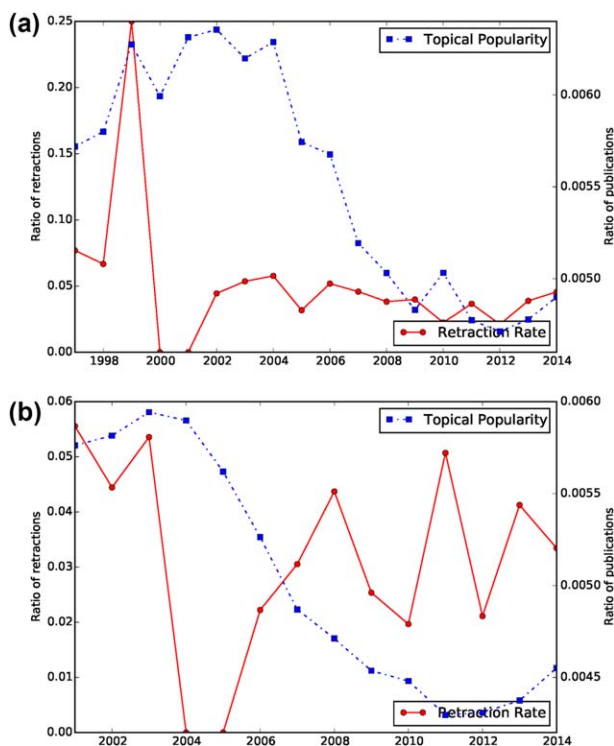


FIG. 4. The trend of topical popularity ($Pop^k(y)$) and retraction rate ($Ret^k(y)$) of two topics: (a) regulation of gene expression and (b) apoptosis. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 4. Method of selecting treatment group and determining retraction year.

Treatment group	Selection criterion	Retraction year
P^t	$P^t = P$	The year of announcement of retraction notice
A^t	The first author from A	The retraction year of the author's first retracted paper
I^t	The institution associated with A^t	The retraction year of the institution's first retracted paper
P_{citing}^t	The earliest paper that cites P^t	The retraction year of the cited retracted paper
P_{coref}^t	The paper that shares the most references with P^t	The retraction year of the co-referring retracted paper
A_{coaut}^t	The most frequently collaborating innocent coauthor with A^t	The retraction year of the collaborating retracted author

Selecting Control Entities

The basic principle of selecting control entities E^c is to ensure the highest degree of similarity between $e^c \in E^c$ and the treatment counterpart $e^t \in E^t$ previous to the retraction year. In so doing, we are able to more confidently attribute the postretraction differences to the occurrence of retraction. Based on the impact curve $C(y)$, we further define a preretraction distance $PreDis$ between a treatment entity e^t and a control entity e^c as follows:

$$PreDis(e^t, e^c) = \|C^{e^t}(y) - C^{e^c}(y)\|_2, y_0 \leq y \leq y_r \quad (1)$$

where y_r is the retraction year. In addition, we require that e^t and e^c share the same starting year y_0 .

In addition to the impact curve and the preretraction distance, there exist some other metadata associated with each publication that can help us to select control entities, such as publication date, journal, and ESI category. Authors and institutions do not have ESI category information, but we can assign each author and institution the most frequently appearing ESI categories from all of their publications. Given a treatment entity e^t , we adapt the method from Lu et al. (2013) to select control entities as follows:

1. Preselect a set of candidate control entities. Specifically, we ensure that the candidate control papers share the same publication date and publication journal as e^t ; and the candidate control authors and institutions share the same ESI category and starting year y_0 of the impact curve of e^t .
2. For all entities in the candidate control set, we pick the top 10 entities with the minimum preretraction distance ($PreDis$) from e^t .
3. From the above selected top 10 minimum-scoring entities, we further pick the two entities, e^{o1} and e^{o2} , whose citation counts before retraction are the closest to e^t , as the final selected control entities for e^t .

Significance Testing

Once the treatment and control groups are determined, we use the following two metrics to compare them: postretraction impact and impact change ratio. Preretractation, or postretraction impact is defined as the sum of citation counts, before or after, the retraction year; impact change ratio is defined as the ratio of postretraction impact to preretraction impact. We test whether the differences for each of the two metrics between the groups are statistically

significant. If so, we claim that retraction does affect the scholarly impact of the involved academic entities. During the process of control entities selection, each treatment entity e^t is associated with two control entities e^{o1} and e^{o2} . Therefore, we compute the mean postretraction impact, or mean impact change ratio, of e^{o1} and e^{o2} , and compare it with the corresponding single value of e^t .

Rather than the more commonly used t -test to compare group means between normally distributed samples, we chose to use the Mann–Whitney U -test (i.e., nonparametric version of t -test) to compare the median values of the treatment and control groups because the distribution of citation counts is highly skewed (see Figure 1(b)).

Comparative Test Results

Table 5 lists the comparative experimental results for retracted and related entities, and their corresponding selected control entities. The results show that the impact of retracted publications and authors significantly decreased after retraction, indicating that the scientific community gradually withdraws academic recognition of retracted papers and authors. Interestingly, the scholarly impact of institutions involved in retractions is consistently higher than that of their control group. We hypothesize that this is because many retracted institutions are well-established institutions with good reputations, like Harvard University and MIT, and that retractions from a specific professor or research group hardly affect the reputation of the whole department or university. Finally, the scholarly impact of papers and authors related to retracted papers or authors seems not to be significantly affected, as long as those related papers and authors are not involved in any retraction cases.

TABLE 5. Results of comparative tests.

Treatment group	No. of samples	Postretraction impact comparison	Impact change ratio comparison
P^t	2659	$2 < 6^{**}$	$0.5 < 1.12^{**}$
A^t	2092	$69.5 < 80.5^{**}$	$0.76 < 0.93^{**}$
I^t	2035	$670 > 375^{**}$	$3.72 > 2.48^{**}$
P_{citing}^t	1889	$5 = 5$	$1.33 > 1.22$
P_{coref}^t	2335	$5 > 4^*$	$1 = 1$
A_{coaut}^t	1915	$285 > 261^*$	$0.62 > 0.60$

Note. In the last two columns, the numbers before and after comparison symbols are median values from treatment groups and control groups, respectively. * p -value less than .05; ** p -value less than .01.

TABLE 6. Segmentation analysis of scholarly impact change ratio as a result of different retraction reasons.

Retracted entity	Overall	Misconduct— media covered	Misconduct— falsification	Misconduct— plagiarism	Misconduct— violation	Error
P^t	0.5	0.23	0.33	0.5	0.5	0.5
A^t	0.76	0.31	0.54	0.77	0.58	0.79

Retraction as a result of falsification or fabrication causes the most serious decrease of scholarly impact, and media coverage further exacerbates this trend.

Segmentation Analysis

Figure 2 shows the distribution of papers by reason for retraction. We also classify scholars and institutions by reasons for retraction based on their retracted papers. In this section, we look at the effect that the reason for retraction itself may have on the scholarly impact of papers and authors. We hypothesize that certain types of reasons for retraction may have more impact than others.

In addition, we want to examine whether the increased attention brought by media coverage further decreases the scholarly impact of papers and authors involved in retractions. To do this, we extracted notable cases of scientific misconduct (most of them are falsification cases) reported by mass media and highlighted in Wikipedia.³ These cases involved a total of 30 scholars and their retracted papers, all of which can be found in our WoS corpus.

Table 6 compares the median values of the impact change ratio for retracted publications and authors, organized by retraction reason. The smaller the value, the greater the decrease in scholarly impact after retraction. We can see that retractions because of falsification or fabrication have the most substantial decrease in their scholarly impact for both papers and authors. This decrease is even more pronounced when the retraction cases are exposed to the public by media.

Comparison With Prior Relevant Studies

Our statistical analysis of the effect on scholarly impact as a result of retraction is built on several prior studies. Here we compare our experimental data, methodology, and results with the key studies that most directly influenced our work.

$$P^t \text{ vs. } P^c$$

Lu et al. (2013) proposed the basic framework of designing comparative statistical experiments to quantify the effect of retraction on scholarly impact (e.g., citation counts), which has been utilized by several follow-up research, including ours. They analyzed a set of articles from the WoS database and found that the citations to retracted articles drop more sharply when compared with nonretracted control articles after retraction, which are quite similar to our study in terms of dataset and results. However, there are three main differences: First, they only used articles published after 2000, whereas we used articles since 1980, and the size

of our data is almost twice as large as theirs; second, they applied a Poisson regression model to quantify the factor of retraction to citation counts in the postretraction period, whereas we used a nonparametric statistical test to compare the median of citation counts between retracted and control articles; third, they divided the retracted articles into two categories, that is, self-reported and nonself-reported, whereas we categorize them into more diverse categories based on different retraction reasons for segmentation analysis.

$$A^t \text{ vs. } A^c$$

Azoulay et al. (2015) investigated the effect of retractions to scholar's reputation, and found that the citation rate to their articles dropped by 10% on average after retraction occurred. Especially, those eminent scientists whose retraction cases involve fraud are more harshly penalized than less-distinguished peers. Their results are also similar to ours, even though there are three main differences: First, their study focused on 878 biomedical research articles from PubMed and 376 U.S.-based retracted faculty authors, whereas our dataset is more broad and larger, including more than 2,000 retracted articles from WoS and their associated authors from all over the world, and covering more than 20 categories of scientific fields; second, they match the names of retracted authors with the Faculty Roster of the Association of American Medical Colleges, whereas we applied a more advanced and patented tool to normalize and disambiguate authors names; third, similar to Lu et al. (2013), they applied Poisson regression analysis to quantify the effect of retraction to an author's citation rate, whereas we used a nonparametric statistical test to compare the median of citation counts between retracted and control authors.

$$P_{coref}^t \text{ vs. } P_{coref}^c$$

Azoulay et al. (2014) studied the effects of retracted articles on their topically related but nonretracted articles. According to their findings, the citation counts to those related articles also declines compared with control articles, which is the opposite of our findings in this study. We may attribute the difference to several aspects. First, we used a different retracted articles dataset. They used retracted articles from PubMed, whereas we used retracted articles from WoS, with around 20% of the articles not indexed by PubMed. Second, we used a different strategy to select "related" articles. They leveraged the overlapping of *MeSH*

³https://en.wikipedia.org/wiki/Scientific_misconduct

keywords between two articles to measure their relatedness, whereas we leveraged the overlapping of reference sections between two articles to measure their relatedness. In addition, they selected the top 100 most related articles for each retracted article (i.e., total of around 60k related articles for 1k retracted articles), whereas we only selected the most related article. The reason why we only selected one related article paired with each retracted article is to avoid the potential misinterpretation of tiny yet significant differences, only because we used a very large dataset. Finally, similar to Lu et al. (2013), they applied Poisson regression analysis to quantify the effect of retraction to the citation rate of related articles, whereas we used a nonparametric statistical test to compare the median of citation counts between retraction-related and control articles.

$$A_{coaut}^t \text{ vs. } A_{coaut}^c$$

Mongeon and Larivière (2015) found that scholars who once coauthored with fraudulent scholars suffered from a decline in their scholarly reputation and productivity, through designing comparative tests on data from WoS. Although we use a similar dataset and methodology, we reach a totally different conclusion. This is not surprising, because their focus is on authors who collaborated with fraudulent scholars on retracted papers, whereas our focus is on authors who collaborated with fraudulent scholars but on nonretracted papers.

In addition to the above pairs of comparative tests for which similar prior studies can be found, we also statistically compared P_{citing}^t vs. P_{citing}^c and I^t vs. I^c . Overall, our experimental results confirmed the previous studies that articles and scholars that are directly involved in retractions will be penalized in terms of their citation rate by academia, through using a larger scope of dataset and different statistical inference methodology. However, we did not find a significant effect of retractions on innocent article and scholars, even if they're indirectly related to those retracted articles or scholars. Especially, we noticed that the study from Azoulay et al. (2014) reached a different conclusion from us on this aspect, by using a different dataset and methodology. We will leave it for future work to further explore and discuss why such differences exist.

Granger-Causality Analysis for Effect of Retraction on Topical Popularity

Ideally, we could apply the same comparative method to investigate the effect of retraction on scientific topics. However, it turns out to be very difficult to find a set of control topics that are comparable to the treatment topics of retracted papers. Instead, we apply the Granger causality test to examine whether the yearly topical retraction rate ($Ret^k(y)$) could statistically cause, or in other words, predict, the value of future yearly topical popularity ($Pop^k(y)$), for the top 10 retracted topics listed in Table 3.

The Granger Causality Test is a statistical hypothesis test for determining whether one time series is useful in

TABLE 7. Granger Causality test of the possibility that retraction rate $Ret^k(y)$ affects topical popularity $Pop^k(y)$.

Topic	$n=1$	$n=2$	$n=3$
Regulation of gene expression	0.4	0.29	0.34
Gene expression	0.57	0.01	0.004
Protein	0.5	0.3	0.22
Apoptosis	0.01	0.16	0.39
Cell (biology)	0.02	0.19	0.12
Gene	0.24	0.17	0.56
Enzyme inhibitor	0.64	0.98	0.99
Cancer	0.15	0.45	0.69
Tumor	0.74	0.66	0.97
Inflammation	0.1	0.65	0.36

forecasting another. Specifically, a time series $X(t)$ is said to Granger-cause $Y(t)$ if it can be shown, usually through a series of t -tests and F -tests, previous $X(t)$ values that provide statistically significant information about future values of $Y(t)$ Granger (1969). As shown in Equation 2, the Granger Causality test can be mathematically formulated as a linear regression model, where $Y(t)$ is the predicted variable, $X(t)$ is the predicting variable, C is a constant, $E(t)$ is the error term, A_i and B_j are corresponding coefficients, and n is the number of lagged years.

$$Y(t) = \sum_{i=1}^n A_i Y(t-i) + \sum_{j=1}^n B_j X(t-j) + C + E(t) \quad (2)$$

In our case, the Granger Causality test can help us to determine whether $Ret^k(y)$ Granger-causes a change of $Pop^k(y)$. In other words, whether retractions in a topic led to an overall decrease in that topic's future popularity.

Table 7 lists the p -values of Granger Causality tests predicting $Pop^k(y)$ in the top 10 most-frequently retracted topics, based on Equation 2 and set $n = 1, 2, 3$, respectively. The rest of the less-frequently occurring retracted topics are ignored, to avoid the effect of the data sparsity problem on the correctness of tests. Most of the test results are not significant (i.e., $p > .05$), except for three topics: gene expression, apoptosis, and cell. After checking the corresponding coefficients of previous popularity $Pop^k(y-i)$ (i.e., A_i), and previous retraction rate $Ret^k(y-j)$ (i.e., B_j) from Table 8, we can see that the magnitude of B_j is close to zero and much smaller than A_i . This suggests that the effect of the historical retraction rate on the topical popularity is almost negligible, compared with that of historical popularity, even if such a tiny effect seems to be statistically significant. Overall, the results from Granger Causality tests show that the retraction rate in one topic hardly affects its future popularity.

Discussion and Conclusion

Some common themes related to retractions, as we raised at the start of the paper, are investigated and discussed: What does a typical retracted paper look like? How does the scientific community react to paper retractions? To what extent does an increase in retractions impact academia and

TABLE 8. Coefficients of Granger Causality tests with significant p values.

Topic	$n = 1$	$n = 2$	$n = 3$
Gene expression	NA	$A_1 = 1.20,$ $A_2 = -0.22;$ $B_1 = -0.003,$ $B_2 = -0.005$	$A_1 = 1.26,$ $A_2 = -0.01,$ $A_3 = -0.33;$ $B_1 = -0.006,$ $B_2 = -0.005,$ $C_2 = 0.002$
Apoptosis	$A_1 = 0.94;$ $B_1 = 0.006$	NA	NA
Cell (biology)	$A_1 = 0.91 ;$ $B_1 = -0.001$	NA	NA

Note. A_i represents the coefficient of $Pop^k(y-i)$ and B_j represents the coefficient of $Ret^k(y-j)$.

science? To investigate the above questions, we conducted a comprehensive study and exploration of a set of retracted articles from Thomson Reuters WoS across 30 years. Our main findings are summarized as follows.

First, papers are retracted for multiple reasons, with the most frequent reason being scientific misconduct. Based on 1,666 annotated retracted articles in our study, scientific misconduct accounted for more than 50% of retractions. The second most common reason for retraction was accidental errors, which comprise around 24% of retractions. Within all retractions because of scientific misconduct, plagiarism occurs most frequently, followed by fabrication and falsification. With regard to the research subject, the top three subjects of retracted publications are molecular biology and genetics, immunology, and microbiology.

Second, the scholarly impact of retracted papers and their authors significantly decreases after retraction. The results of our comparative experiments show that retracted papers receive significantly fewer citations after retraction than control articles published in the same journal on the same date. In addition, the reputation of scholars listed as first authors on retracted articles decreases significantly in the postretraction period compared with other scholars having similar reputation growth patterns and research focus before retraction. In particular, those scholars who have falsified or fabricated data in their research received an even more severe decrease in their reputation, and media coverage further aggravates this penalty. Decreased scholarly recognition, based on citation count, of retracted papers and scholars, reflects the punitive reaction of the scientific community to falsified or erroneous scientific studies.

Third, even when retracted papers and their authors are penalized by the scientific community with fewer citations, this negative effect was not shown to propagate to their sponsoring institution, or to other related but innocent papers and scholars. According to our comparative experimental results, the reputation of those institutions that sponsored the scholars who were accused of scientific misconduct did not seem to be tarnished at all. In addition, neither was a

propagation effect observed for papers topically related to retracted articles based on citation relationship, nor for those scholars related to retracted authors based on coauthorship. Moreover, the Granger Causality analysis of dynamic retraction rate to the overall popularity of a certain scientific topic shows that the phenomenon of retraction bears no apparent effect on the popularity of a research topic. All these results strongly suggest that the penalty effect of paper retractions is quite localized to the authors who are mainly responsible for the retraction and does not extend to the wider scientific community.

Scholarly impact is only one aspect that is affected by retraction; there may exist other issues that are also influenced by retraction to be explored in future work. The retracted study from Wakefield et al. (1998) shows that the effects of paper retraction may go beyond academia and extend to other aspects of society. In addition, we find in this study that citing retracted papers does not seem to lead to any serious academic consequences. We cannot provide details as to why this is the case, but hypothesize that full-text citation analysis may help us better understand the contextual information of where and how retracted papers were cited. By further analyzing the text surrounding the citations to retracted articles, it might be possible to understand the author's attitude and intention in citing the retracted study, and identify any possible negative influence as a result of the citing behavior. For instance, van der Vet and Nijveen (2016) conducted a full-text analysis toward the whole citation network of one typical retracted article published in *Nature*, and showed that the retracted results contained in the original retracted article may propagate to its directly citing articles. However, they did not provide evidence that citing retracted articles will also lead to errors in their own studies, which can be an interesting research problem for our future work.

Acknowledgments

We thank our colleagues William Dowling, who provided the Web of Science dataset and tools for author name disambiguation, as well as James Pringle for insightful comments on this article.

References

- Atlas, M.C. (2004). Retraction policies of high-impact biomedical journals. *Journal of the Medical Library Association*, 92, 242–250.
- Azoulay, P., Bonatti, A., & Krieger, J.L. (2015). The Career Effects of Scandal: Evidence from Scientific Retractions. Working Paper 21146 National Bureau of Economic Research.
- Azoulay, P., Furman, J.L., Krieger, J.L., & Murray, F. (2014). Retractions. *Review of Economics and Statistics*, 97, 1118–1136.
- Bornemann-Cimenti, H., Szilagyi, I.S., & Sandner-Kiesling, A. (2016). Perpetuation of retracted publications using the example of the Scott S. Reuben case: Incidences, reasons and possible improvements. *Science and Engineering Ethics*, 22, 1063–1072.
- Chen, C., Hu, Z., Milbank, J., & Schultz, T. (2013). A visual analytic study of retracted articles in scientific literature. *Journal of the Association for Information Science and Technology*, 64, 234–253.
- Elizabeth, W. (2015). Why are retractions so difficult? *Journal of Science Education*, 2, 32–34.

- Eridani, S., Sgaramella, V., & Cova, L. (2004). Stem cells: From embryology to cellular therapy? An appraisal of the present state of art. *Cytotechnology*, 44, 125–141.
- Fanelli, D. (2013). Why growing retractions are (mostly) a good sign. *PLoS Medicine*, 10, 1–6.
- Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. *PLoS One*, 10, e0127556.
- Fang, F.C., Steen, R.G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 17028–17033.
- Furman, J.L., Jensen, K., & Murray, F. (2012). Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine. *Research Policy*, 41, 276–290.
- Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438.
- Grieneisen, M.L., & Zhang, M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *PLoS One*, 7, e44118.
- Griffith, R. (2015). Method and system for disambiguating informational objects. US Patent, 9,183,290.
- Hwang, W., Roh, S., Lee, B., Kang, S., Kwon, D., Kim, S., ... Schatten, G. (2005). Patient-specific embryonic stem cells derived from human scnt blastocysts. *Science*, 308, 1777–1783.
- Hwang, W.S., Ryu, Y.J., Park, J.H., Park, E.S., & Lee, E.G., (2004). Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst. *Science*, 303, 1669–1674.
- Intellectual Property & Science. (2013). Data collection, cleaning and processing. Technical Report Thomson Reuters.
- LaCour, M.J., & Green, D.P. (2014). When contact changes minds: An experiment on transmission of support for gay equality. *Science*, 346, 1366–1369.
- Lu, S.F., Jin, G.Z., Uzzi, B., & Jones, B. (2013). The retraction penalty: Evidence from the Web of Science. *Scientific Reports*, 3, 3146.
- Madlock-Brown, C.R., & Eichmann, D. (2015). The (lack of) impact of retraction on citation networks. *Science and Engineering Ethics*, 21, 127–137.
- Michalek, A.M., Hutson, A.D., Wicher, C.P., & Trump, D.L. (2010). The costs and underappreciated consequences of research misconduct: A case study. *PLoS Medicine*, 7, e1000318.
- Mongeon, P., & Larivière, V. (2015). Costly collaborations: The impact of scientific fraud on coauthors' careers. *Journal of the Association for Information Science and Technology*, 67, 535–542.
- Noorden, R.V. (2011). Science publishing: The trouble with retractions. *Nature*, 478, 26–28.
- Pfeifer, M.P., & Snodgrass, G.L. (1990). The continued use of retracted, invalid scientific literature. *Journal of the American Medical Association*, 263, 1420–1423.
- Resnik, D.B., Wager, E., & Kissling, G.E. (2015). Retraction policies of top scientific journals ranked by impact factor. *Journal of the Medical Library Association*, 103, 136–139.
- Steen, R.G., Casadevall, A., & Fang, F.C. (2013). Why has the number of scientific retractions increased? *PLoS One*, 8, e68397.
- Tabar, V., & Studer, L. (2002). Novel sources of stem cells for brain repair. *Clinical Neuroscience Research*, 2, 2–10.
- Tör, M., Brown, D., Cooper, A., Woods-Tör, A., Sjölander, K., Jones, J.D., & Holub, E.B. (2004). Arabidopsis downy mildew resistance gene rpp27 encodes a receptor-like protein similar to clavata2 and tomato cf-9. *Plant Physiology*, 135, 1100–1112.
- van der Vet, P.E., & Nijveen, H. (2016). Propagation of errors in citation networks: a study involving the entire citation network of a widely cited paper published in, and later retracted from, the journal *Nature*. *Research Integrity and Peer Review*, 1, 10.
- Wakefield, A.J., Murch, S.H., Anthony, A., Linnell, J., Casson, D.M., Malik, M., ... Walker-Smith, J.A. (1998). Ileal-lymphoid-nodular hyperplasia, nonspecific colitis, and pervasive developmental disorder in children. *Lancet*, 351, 637–641.