

MDSAA

Master Degree Program in
Data Science and Advanced Analytics

The concern for literature self-correction in biomedicine

A bibliometric analysis of retractions

Isabel Maria Couto Dias

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

The concern for literature self-correction in biomedicine

A bibliometric analysis of retractions

by

Isabel Maria Couto Dias

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Bruno Damásio, PhD, Nova IMS

Sandro Mendonça, PhD, ISCTE

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 15th of July 2024

ACKNOWLEDGEMENTS

Firstly, I thank my supervisors, Professor Bruno Damásio and Professor Sandro Mendonça, for the encouragement, enthusiasm and guidance. Your help has been instrumental in the completion of this thesis.

I want to express my deepest appreciation for my family. To my parents, Carlos and Maria Helena, and to my brothers, Artur and André, thank you for your advice and especially the unwavering support you have provided throughout this journey.

I am also grateful to the friends I made during my degree. Our brainstorming sessions, shared ideas, and collaborative problem-solving made the process more enjoyable and creative. To the friends who supported me outside the university, thank you for giving me strength and encouragement.

ABSTRACT

The number of retractions issued every year has shown an increasing trend, a phenomenon that has raised questions regarding scientists' research integrity and the mechanisms set in place to identify potential problems. The biomedical area, in particular, has been a source of a high rate of retractions. This study aimed to determine whether the retractions reflect heightened concern in post-publication corrections to the scientific literature or the consequence of a poor self-policing system. Several indicators suggest that the surge in retractions can be attributed partially to increased scrutiny. While more journals are retracting articles, the retraction rate per journal remained relatively stable until 2020. Articles are faster to retract, while at the same time, articles further in the past are also being removed more. Moreover, citations to retracted articles have an immediate and prolonged decline. Given this issue's complexity, it is impossible to conclude that concerns are more pronounced in biomedicine compared to other fields, as various analyses reveal divergent results regarding which scientific domains exhibit the highest levels of concern. However, this analysis provided a comprehensive view of various measures of concern.

KEYWORDS

Retracted articles; Biomedicine; Research Misconduct; Retraction Citations; Bibliometrics

Sustainable Development Goals (SDG):



TABLE OF CONTENTS

1. Introduction.....	8
2. Literature review	11
2.1. Key Theoretical Concepts	11
2.2. Retractions in Academic Publishing	12
2.2.1. Impacts and Characteristics	12
2.2.2. Literature on Correction Efforts	14
2.2.2.1. Time-to-retraction	14
2.2.2.2. Post-retraction citations	16
2.2.2.3. Pre-retraction citations	18
2.3. Limitations in Existing Literature and How to Address Them	18
3. Materials and Methods	20
3.1. Data Sources, Tables and Pre-processing.....	20
3.2. Analytical Methods.....	24
4. Results and discussion	27
4.1. Retraction Characteristics	27
4.2. Time-to-retraction	31
4.3. Post-Retracton Citations.....	35
4.4. Corrections to Pre-Retracton Citations	39
4.5. Holistic Variable Assessment.....	39
5. Conclusions.....	42
Bibliographical References	44
Appendix A: Classification of Variables	52
A1. Biomedicine.....	52
A2. Misconduct.....	53
Appendix B: Additional Results	56
Annex A: Survival Models.....	62
Annex B: Negative Binomial Regression	65
Annex C: Proportional Stratified Random Sampling	66
Annex D: Cluster Analysis.....	67

LIST OF FIGURES

Figure 3-1. Methodological approach	22
Figure 3-2. Evolution of number of article retractions	24
Figure 4-1. Scientific Area combinations.....	27
Figure 4-2. Evolution of number of retractions and journals according to scientific area	28
Figure 4-3. Rise in number of retractions per year	28
Figure 4-4. Reasons for retraction as a proportion of total retraction articles	29
Figure 4-5. Evolution of misconduct	30
Figure 4-6. Retractions in groups of countries.....	31
Figure 4-7. Kaplan-Meier survival curve of retracted articles.....	32
Figure 4-8. Evolution of time to retraction by publishing year.....	34
Figure 4-9. Evolution of time to retraction by retraction year	34
Figure 4-10. Top cited articles post-retraction	36
Figure 4-11. Evolution of citations in relation to retraction year	37
Figure 4-12. Misconduct vs. Biomedicine number of citations	37

LIST OF TABLES

Table 3-1. Descriptive statistics numerical key variables.....	23
Table 3-2. Descriptive statistics categorical key variables	23
Table 4-1. Time to retraction descriptive statistics.....	31
Table 4-2. Average time to retraction (in months) by reason	32
Table 4-3. Cox-Hazard Ratio Model Summary	33
Table 4-4. Yearly citations regression time evolution.....	38
Table 4-5. Yearly citations regression of comparing pre and post retractions.....	39
Table 4-6. Clusters average per variable	40

LIST OF ABBREVIATIONS AND ACRONYMS

APC	Article Processing Charges
AVG	Average
BIOMED	Biomedicine
COEF	Coefficients
COPE	Committee on Publication Ethics
CRPI	Public Concerns Regarding Publication Integrity
DOI	Digital object Identifier
HR	Hazard Ratios
MAX	Maximum
MED	Median
MIN	Minimum
MMR	Measles, Mumps and Rubella
RWD	Retraction Watch Database
SCI	Scimago
TTR	Time to retraction
WoS	Web of Science

1. INTRODUCTION

The progression of scientific advancements relies on mutual trust in the integrity and rigor of the existing scientific literature. The growing pressure to discover new and groundbreaking ideas, coupled with the “publish or perish” culture, has led to temptation of unethical behaviour for financial and career advancement (Kurt, 2018). Additionally, the journals themselves also have reasons for unethical practices, with the spread of the Open Access Publishing and Article Processing Charges (APC) accepting any article to receive their corresponding fees (Heen & Vogt, 2024).

For this reason, it is imperative that there are corrective measures set in place. Depending on the identified problem and it's the severity, published articles can either be corrected or retracted. According to the Committee on Publication Ethics guidelines (COPE Council, 2019), an article should be retracted in cases where there are major flaws that cannot be remedied through a correction, either as a result of unethical research practices, honest errors or other issues. And even though this mechanism is imperative to maintain the standards of quality in published science and prevent the spread of misinformation, there is evidence that the rate of retraction still does not match the estimates for the number of papers that meet at least one of the criteria for removal (Oransky, 2022). Furthermore, sometimes the decision is not as clear cut as one might expect. For example, a study exploring the support of scientists to the decisions to retract several articles (Sacco et al., 2024) found that, for misconduct or honest errors, scientists were more likely to agree with the decision than when sound scientific findings with potentially negative implications were retracted (information hazards).

Nonetheless, the number of retractions has been on the rise, from 62 retractions in 1997, to 419 in 2007 and a striking 946 in 2014 (Brainard et al., 2018). It is progressively more important to investigate the articles in question and quantify their repercussions. Retractions are often regarded as one of the most stringent forms of self-correction in academia, rendering an entire paper and its findings invalid (Kühberger et al., 2022). Its consequences extend beyond the article itself; it is damaging to everyone involved. Especially in cases of misconduct, the retraction not only affects the paper but also stains the reputation of the researchers, the authors' institutions, and the journal. For the authors, this can lead to a loss of peer acknowledgement in the form of citations or future collaborations, obstacles in future publishing, and even job loss (Azoulay et al., 2017).

Additionally, the flawed knowledge in the retracted articles can result in ill-informed policy decisions and other actions that can significantly affect the lives and well-being of individuals (Rosenstock & Lee, 2002). One of the most infamous cases of misconduct in a publication was by Wakefield et al. (1998) that correlated the measles, mumps and rubella (MMR) vaccines to autism in children. Although multiple studies have since been published disproving these conclusions, leading to its ultimate retraction in 2010, due to the wide publicity the retracted paper received, the MMR vaccination rates decreased significantly, impacting the lives of the

countless unvaccinated children and the vulnerable people around them (Rao & Andrade, 2011). The field of biomedicine may be particularly susceptible to motivating scientists into unethical behaviour due to the association with commercial opportunities (Ioannidis, 2012) and access to grants. It is also an area where unreliable results can put patients at risk, both directly like the case of Wakefield, or indirectly, by the ghost-writing of pharmaceutical companies of articles advocating for certain drugs (Sismondo, 2008).

Given these far-reaching implications, the research field of scientific misconduct has tried to explore the factors that lead to misbehaviour and the quantification of impact. The analysis of the former promotes the transparency and integrity in scientific research, by showcasing instances of misconduct, and encourages researchers to report errors and misconduct (Marcus & Oransky, 2014). The latter is important to assess the resulting damage, to quantify the legal and ethical accountability, and to help rebuild trust.

Unfortunately, the current literature often relies on the analysis of specific authors (Marcovitch, 2007; Teixeira da Silva & Dobránszki, 2017), scientific areas (Ajiferuke & Adekannbi, 2020; Furman et al., 2012; Shepperd & Yousefi, 2023), or geographical areas (Liu & Lei, 2021; Mallapaty, 2024; Park et al., 2018), which often lacks a broader perspective that can be gained by studying comprehensive set of cases. In the search for retractions' impact, many articles limit their scope to the number of citations (Khademizadeh et al., 2023) or only include the time factor into their analysis.

In an endeavour to contribute to the measurement and analysis of research misconduct (Biagioli et al., 2019), this study will quantify the evolution and patterns of the current retractions landscape. More specifically, it was noticed that the biomedical field represents a disproportionate amount of literature in the study of retractions (Kohl & Faggion, 2022), which might be a reflection of a genuine difference in concern towards upholding the standards of scientific literature, or by other factors such as data availability. This prompts the question: Does the biomedical field exhibit a heightened concern towards correcting literature retracted papers compared to other scientific domains? The question will be broken down into two main objectives:

- (i) Identify temporal, geographical, journal trends of papers citing retractions, according to their scientific area;
- (ii) Quantify efforts to correct citations of retractions.

The first objective is essential in measuring differences between the behaviour of the two grouped scientific areas. The second objective is at the crux of the matter: it will allow us to identify if the behaviour towards biomedical literature is different than in other areas. More specifically, the concern will be measured by three distinct factors: the time to retraction, the post-retraction citations the article receive, and the number of pre-retraction citations that emitted corrections after the retraction was issued. We consider that these factors represent the concern of scientists over the different stages of the retraction cycle.

This study will be covering the years 2000 to 2022, as we are more interested in understanding the current landscape and impact of the retractions in academic literature. The relevant data will be sourced from multiple publicly available databases, namely the *Retraction Watch Database* (RWD), *Web of Science* (WoS), and *Scimago* (SCI). Each of them has an important role in assuring the quality of the analysis, even if the conjugation of these sources may lead to a smaller number of overall records in the end.

Finally, this thesis will be divided into the following sections: the Literature Review will provide an overview of what has been done in this field of study and detail the necessary background to understand the subsequent analysis. The Methodology will delineate the steps taken to obtain the results, to guarantee the reproducibility of this study. In the Results and Discussion section, the discoveries we made will be shown and compared to previous literature. Finally, the Conclusions will encapsulate all the relevant findings and their implications.

2. LITERATURE REVIEW

This section aims to detail the current literature available, addressing the objectives outlined in the introduction, and provide the necessary background for the understanding of the following sections. More specifically, this review will identify the works on temporal, geographical, and journal trends of papers citing retractions (Objective 1); and quantify efforts to correct citations of retractions (Objective 2).

2.1. KEY THEORETICAL CONCEPTS

In this study, certain concepts are closely connected, making it essential to clearly define each one. This section contains a concise explanation of each concept, along with any decisions on notation, will be given within the context of this paper.

The retraction of an article, or retraction notice, is an official statement by a journal or publishing entity that formally announces and explains the reasons for the retraction of a previously published article. A retraction is issued when a published article contains significant flaws that cannot be rectified through a correction. The Retraction Guidelines by the Committee on Publication Ethics (COPE) outlines criteria warranting retractions, such as misconduct (e.g., falsification of results, plagiarism, redundant publication), serious inadvertent errors that invalidate the presented conclusions, or legal issues. The article under these conditions is labelled as retracted, but only in very limited cases should the article be removed from the journal, retaining the metadata (title and authors) of the article to keep record.

A similar concept is “withdrawal”, which takes place when an article, either before or after undergoing the peer review stage, is removed from consideration (Charlesworth Author Services, 2022). This contrasts with retractions, as the article is prevented from being published in the journal altogether.

Before an article is retracted, it is also possible to issue an expression of concern. It signals to the viewers that, while there might not yet be conclusive evidence of major errors, the article contains concerns that are sufficiently serious to warrant a warning to potential readers. These notices may be temporary or permanent. In the first case, a further notice is issued to update the reader on the matter, either to a permanent expression of concern, a retraction, or an editor’s notice of exoneration, outlining the outcome of the investigation (Elsevier, 2024a).

Another path that will be explored regards corrections. When minor changes are necessary to correct a published publication, the author can issue a corrigendum in which they detail the problems identified. If the error was made by the publisher, the correction is named an erratum (Elsevier, 2024b).

2.2. RETRACTIONS IN ACADEMIC PUBLISHING

2.2.1. Impacts and Characteristics

When errors or misconduct infiltrate the landscape of scholarly discourse, it can have far-reaching implications that may extend beyond the scientific community, including public perception, future research and even policy making. In the medical field, there have been several instances where the malpractice has led to the potential harm of patients. Steen (2011a) studied the degree to which patients were put at risk, by analysing retracted clinical studies and obtaining all related articles. It was found that 94% of citations to the retracted articles in the study were research related, and only 7% were related to the reason to retraction. This not only indicates that the information could be propagated to other research, but also that citing authors probably were not aware of the document status. From the retracted papers, 9189 patients were treated, and from the secondary articles, 70501 patients were treated. Additionally, the papers retracted for fraud treated significantly more patients, and appeared more often in literature.

Regarding the dissemination into policy making, Malkov et al. (2023) findings indicate that 2.3% of retracted research is cited in literature relevant to policy-making—a rate similar to that of citations for non-retracted articles in policy literature. It was observed that several reasons for retraction, especially those stemming from errors, are more frequently cited than others, regardless of whether academic misconduct is involved. Additionally, although negative citations exist, they represent a minority of the overall citations.

In the business world, Freilich & Kim (2023) analysed to what extent did research misconduct affect the patents system. They found that the patent system was not sensitive to information quality. From the perspective of the applicants, it was shown the continuous prosecution and maintenance of unsupported patents, possibly because of a disconnect between the institutions responsible for maintaining the patents and the news of retraction. For those applicants, whose patent was rejected due to citations of prior unsupported works, the probable explanation for its use most likely resides in lack of awareness. The same reason applies to the citations of retraction papers by examiners.

Measuring the integrity of the scientific community is a complex problem, because it is not possible to quantify it directly. Instead, one must use proxies like surveys (Martinson et al., 2005), investigation reports (Fanelli, 2013) or retractions (Nath et al., 2006). Focusing on the latter, the use of this mechanism started to become more widely used, accompanied by an increase in studies of the behaviour and characteristics of such self-correcting mechanisms (Marcus & Oransky, 2014). Additionally, Steen et al. (2013) found that articles were being removed at an increasing speed, which he correlated with the higher levels of supervision. Overall, it was concluded that, while more scrutiny had some significance in the time to

retraction, the increase in retracted articles reflected the changes in behaviour of both authors and institutions. The rise in both the number and proportion of retractions among authors with a single retraction are supported by the reduced barriers to the publication of flawed research. On the other hand, lower barriers of what constitutes as retraction-worthy are apparent through the new types of reasons for retraction.

Likewise, Fanelli (2013) and Oransky & Marcus (2016) argue that the surge in retractions is a consequence of the emergence of new tools facilitating misconduct detection, a similar evolution of tools that enable deceit. Additionally, the evolution of detection systems, coupled with the increased attention and engagement of scientists in their research areas, contributes to this trend. Despite the rising number of queries to the US Office of Research Integrity (ORI), there has not been a corresponding increase in the frequency of detected misconduct. Moreover, the retractions as a mechanism are still in expansion, as evidenced by the growing number of journals that have started issuing retractions, also confirmed by Vuong et al. (2020).

Furman et al. (2012) found that articles who receive more attention (in the form of citations) are more likely to be retracted due to the effect of post-publication peer review. It was found that the connection of the corresponding author with a US-Top25 university and the number of citations in the first year of publication were two characteristics that increased the likelihood of retraction. On the other hand, Tang et al. (2020) found that retracted papers predominantly involved multiple authors, and repeat offenders are more inclined to collaborate with others. Yet, the study also revealed that collaboration alone did not increase the likelihood of generating research warranting a retraction.

Fanelli et al. (2015) refuted the belief that factors like gender or high pressure to publish played a role in propensity for retraction. On the other hand, it found that countries lacking research integrity policies, where individual publications were monetarily rewarded, cultures and situations that hindered mutual criticism, and authors' early careers were found to be better predictors of misconduct.

A substantial share of articles uses *PubMed* as their source of bibliometric information, which restricts their scope to biomedical articles. Such is the case with Fang et al. (2012), that found misconduct accounted for most of the retractions of scientific publications, while only 21,3% were attributable to error. That information contrasts with previous literature, since many retraction reasons were re-labelled to fraud, by taking into account secondary sources of data besides the retraction notices, like the ORI reports. It was also found that the rise in retractions was not uniform across the different reasons, as fraud or suspected fraud increased at a higher rate than errors. From the geographic information, it was concluded that the United States, Germany, Japan and China accounted for 75% of the retractions due to fraud or suspected fraud.

Other scientific areas like Computer Science have also started to delve into their own analysis of the phenomena. One example is Shepperd & Yousefi (2023), who found that in comparison

to other fields, the 56% retracted papers provided scarce or no information about the reason for retraction, varying widely between publishers. There is also evidence of multiple copies of the article that may not include a notice of retraction, leading to post-retraction citations.

In the Engineering Field, Rubbo et al. (2019) used the *Web of Science* Database to study a total of 238 retractions published by 117 journals. They noted that the first recorded retraction was in 2002, but only in 2008 did they start to become more common. Analysing geographic tendencies, the results showed that articles published in England were leading the way, followed by the United States and the Netherlands. So, unsurprisingly, the continent that presented the most retractions was Europe, followed by North America and Asia.

With the emergence of tools like the *Retraction Watch Database* (The Center for Scientific Integrity, 2018), studies were able to broaden their scope. Brainard et al. (2018) based their study on this data and found that only 500 authors were responsible for about one-quarter of all retractions analysed. It also found that countries with smaller scientific communities rank higher in rate of retraction. Similarly to other studies, it also found the rise of academic misconduct, in various forms, and the correlation between journals with higher impact factor and number of retractions.

2.2.2. Literature on Correction Efforts

This study will try to quantify the concern towards erroneous scientific literature in three different fronts: the time-to-retraction of each area, the post-retraction citations the retracted articles receive, and the pre-retraction citations that, after retraction, had some kind of correction associated with it.

2.2.2.1. Time-to-retraction

Within the three research lines described above, multiple studies have explored the role of time-to-retraction. In a study comparing medical genetics articles with non-medical genetics articles, Dal-Ré & Ayuso (2019) discovered that when plagiarism or fabrication/falsification was involved, the more recent articles were retracted in a shorter time span. Additionally, it was found that articles involving fabrication/falsification were retracted more quickly in medical genetics articles compared to non-medical genetics ones. However, this discrepancy was not observed in articles retracted due to plagiarism, except during the period from 2016 to 2018. Contrary to these findings, investigations conducted by the ORI, primarily focused on articles retracted for fabrication/falsification, concluded that medical genetics articles and non-medical genetics articles did not exhibit a statistically significant difference in time-to-retraction, with means of 4.6 and 5.1 years, respectively.

Nath et al. (2006) conducted a study on retracted articles indexed in MEDLINE and discovered that articles retracted due to unintentional errors were more likely to be removed by the authors themselves and had a prompt response time. On average, these articles had a shorter retraction time compared to those retracted for misconduct, with an average of 2 years versus 3.3 years, respectively.

Additionally, Redman et al. (2008) took retraction data from *PubMed*, spanning from 1995 to 2004, and found that, on average, it took 20.75 months between publication and retraction, but this interval decreased over time. This value contrasts with the 28 months' time lag between publication and retraction that Budd et al. (1998) measured, with data from 1966 to 1994. This suggests an evolution of the proactiveness towards correcting literature.

Steen (2011b) analysed the time to retraction of the retracted articles from *PubMed* separated by publication type (fraudulent papers vs. scientific mistake). The months from publication to retraction has increased over the years, for both publication types. Moreover, a correlation between total retractions (per year) and the time to retraction was noticed as well as a correlation between time of retraction of fraudulent and mistake papers. It is attributed to journals reaching further back to retract published papers. Among papers retracted for fraud, in 2000, the longest time to retract was 8 months; in opposition to the maximum of 108 months to retraction in 2009.

The most in-depth study about time-to-retraction, to the best of our knowledge, concluded that there was no difference between falsified and non-fraudulent papers in the 21 most high impact journals (Trikalinos et al., 2008). Using Kaplan-Meier plots of the time from publication to retraction, they evaluated if there was a correlation between the time to retraction and number of authors, rank, authorship position of the main implicated researcher, scientific field, country of correspondence, multinational collaboration, funding and citations received within 12 and 12-36 months after publication. The median survival time was 28 months until retraction, with 81% surviving a year or more. The time to retraction was very correlated with the authorship position and the rank, where senior rank and last author were significantly correlated. For articles that were retracted very soon, the rank or author did not matter as much as after the first year: when main implicated author was senior researcher, it took much longer to retract.

When comparing articles retracted for fraud and other reasons, it was concluded that there was no statistically significant difference between time to retraction of top-cited journals and non-top-cited journals for both senior and junior researchers. Comparing Europe to United States and other countries, the difference in time to retraction did not reach statistical significance. The same applies to the number of authors, multinational articles, and number of citations received in the first year. For the articles that received more citations between 12 and 36 months, these were slower to be retracted. Within the fraudulent authors, the time to the first retraction was significantly shorter for junior researchers, where the median for juniors was 23 months and for seniors it was 77 months.

Fang et al. (2012) also studied the time to retraction and found that it varied according to the cause to retraction, where articles retracted for fraud took substantially longer. A gradual increase in time to retraction was also found. Journal impact factor did not correlate with time to retraction except for fraud cases.

2.2.2.2. Post-retraction citations

Another way to measure the “concern” towards upkeeping the scientific record is the number and evolution of post-retraction citations. COPE advises to abstain from citing retracted articles, as they are viewed as invalid to the scientific community. However, it is possible that, even after retraction, the article is disseminated through citations, or other mediums (Khademizadeh et al., 2023). The analysis of citations is interesting to measure several facets of this phenomenon: post-retraction citations indicate whether the retraction has effectively removed the paper from valid literature. It may also reflect the unwillingness of authors to accept the decision to retract their papers, or even get a measure of the caution of the scientists to avoid using bad science to support their work.

Interestingly, there is no consensus in the description of the behaviour of citations after retraction. Furman et al. (2012), notes that following the retractions, future citations have an immediate, severe and long-lived decline. This study used the *PubMed* Database to identify retractions and the *Web of Science* database to obtain the relevant citation information. The follow-up citations were contrasted with citation patterns of non-retracted papers, using a matched control set. It was composed of the nearest neighbours of the retracted articles, i.e. the article before and after of the same journal.

It employed a difference-in-differences analysis on the post-retraction citations, to assess how the rate of citation changes after retraction. It found a significant drop in the citations after retraction. But that drop had no correlation with other factors like the authors being from the US or from a Top 25 university. However, it was noticed that the effectiveness of retraction improved in more recent years, with a 69% decline in retractions since 2000. In fact, it argues that the effect is immediate and severe, observing a more than 50% decline by the second year and 72% by the tenth year.

Another study that corroborates the hypothesis that citations plunge after retraction is Dinh et al. (2019). This study uses a similar data approach, using *PubMed*, the retracted articles were retrieved and matched with Scopus to obtain the citations data. Additionally, this study also included data from the *Retraction Watch Database* to complement with retraction data. An important distinction made by this article is that post-retraction citations were considered as those published at least two years after retraction, to account for the delayed nature of journal publication and retraction process. Contrasting with the previously mentioned study, this found a drop in 86.5% in citations after retraction. Moreover, geographic characteristics and research areas of citations were found to differ from those of the original retracted article.

Using a one-tailed Wilcoxon signed-rank test for non-normally distributed data, it was found that the post-retraction citation counts were significantly lower than the pre-retraction citations. It went further by using ego network analysis to examine the structural changes in each network. By comparing to non-retracted articles to account for the natural citation decay, it was concluded that the drop was indeed a consequence of retraction.

On the other side of the spectrum, Candal-Pedreira et al. (2020) found that the retraction of the articles had no impact on the number of citations across journal quartiles. The articles analysed were those who were published in MEDLINE indexed journals from 2013 and 2015 and retracted between 2014 and 2016 and the corresponding citations those articles received. Its innovation comes from an analysis of a mirroring timespan of the pre and post retraction period. From a statistical descriptive analysis, they found a reduction in the post-retraction citations only in the first quartile journals, with the second and fourth quartile showing an increase in the number of citations.

Heibi & Peroni (2022) drew the same conclusion: the citations do not decrease after the year of retraction. It used data from the humanities domain, a measured quantitative and qualitative properties of the citations of the retracted papers. The analysis of the sentiment behind the citations allowed for the discovery that only few articles mentioned the retraction or were mentioned in a negative light. However, it was noticed that in the medical field, the citations had a higher concern or awareness for retractions, in comparison with the humanities and social science domains.

There are several articles that find reasons for this phenomenon. For example, Schmidt (2018) found the *Web of Science* Database was very incomplete in their labelling of retractions in comparison to the *PubMed* database, which can lead to the usage of articles that are invalid. This is confirmed by Park et al. (2018), who found that only about 50% of the retracted articles are indexed as such in the *Web of Science* Database for retractions of Korean authors compared to the *Retraction Watch Database*. The research platforms also take time to update their information. For example, the number of retraction notices indexed in *PubMed* only stabilized 35 months after the retraction was published (Decullier et al., 2014). In other words, after the article is retracted in the press, there is a time lag until that change is documented in *PubMed*.

Adding on to this, it is common to have several versions of the same paper published in different websites which proliferate the untrustworthy findings (Davis, 2012), and not all be updated with the retraction notice (Shepperd & Yousefi, 2023). These two factors support the hypothesis that the citations of retracted papers result from lack of awareness. Indeed, the majority of papers do not acknowledge the status of the article, with only about 5% doing so (Neale et al., 2010).

2.2.2.3. Pre-retraction citations

In the study of citations to retractions, the focus is mostly on the post-retraction citations, as they are the most direct measure of effectiveness. Unfortunately, current literature either uses the pre-retraction citations as a comparison point to the post-retraction citation pattern or considers all citations as a whole. This means that the information about this topic is very limited and surface level, as its analysis was used as a tool to support the investigation of another topic. However, the pre-retraction citations also can provide valuable insights into the impact of retractions.

For example, Furman et al. (2012) noticed that in years prior to the retractions, there was no decline in citations. This indicates that the retractions were not anticipated by the scientific community. Additionally, using 20 of the retracted papers in the study, the citations they received were analysed according to their sentiment, separated into two categories: retracted before and after 24 months. It was found that only one of the articles retracted in the first 24 months of publication received negative citations, debating the results.

Another paper found that higher attention to articles correlated with a faster retraction (Marcus & Oransky, 2014). The natural conclusion is that there is a “post-publication” peer review process, in which scientists validate the findings in published papers.

2.3. LIMITATIONS IN EXISTING LITERATURE AND HOW TO ADDRESS THEM

The previous section outlined several research lines within this area of study, which shows that there is substantial work written on the retractions themselves and their impact. However, there is room for a better understanding of the effects of this phenomenon. It is very common that articles focus on a specific scientific area (Bhatt, 2021; Eldakar & Shehata, 2023; Khademizadeh et al., 2023; Zheng & Fu, 2022), and explore the problem of retractions through the lens of the particularities of said area. It is especially prevalent that studies focus on the biomedical area, sometimes implicitly, by the sole use of the database *PubMed* (Hsiao & Schneider, 2022; Neale et al., 2010). The work done to capture the impact of the retractions has been either surface level, or analyses very specific cases, unable to capture global trends or patterns.

Additionally, very few studies have looked at the efforts to rectify the literature in response to retractions, and their effects on cited references. Existing articles predominantly emphasize post-retraction citations, where it is assumed that researchers should be aware of the article's status and may not need corrections due to shifts in the literature. Or, alternatively, the pre-retraction citations are used in comparison and as a tool to understand the post-retraction citations. However, there is very limited research on the effect of the pre-retraction citations alone. Examining corrections associated with pre-retraction citations can offer insights into

the extent to which scientists prioritize maintaining the quality of previously published records.

This study distinguishes itself from previous literature in three major ways. Firstly, it broadens the scope of the analysis by including all available retracted articles, only restricting by year. It also compares different scientific areas as a whole instead of analysing a field on its own. Finally, it includes the novel layer of analysis of corrective actions on pre-retraction citations, opening a new avenue for measuring concern towards upkeeping the quality of the scientific record. It will take inspiration from previously applied methodologies and complete it by drawing together the conclusions of each area of analysis.

Past studies of retraction characteristics have heavily relied on the reporting of descriptive statistics and statistical testing (Shema et al., 2019; Trikalinos et al., 2008). Visualisation tools are also an integral part in the display of patterns and to guide the reader through the direction of the analysis and hypothesis tested (Chen et al., 2013; Heibi & Peroni, 2022).

Additionally, this study will harness the computational power of clustering algorithms. They have been used in a myriad of bibliometric studies to find groups of units with similar characteristics. A popular approach is to use the keywords of the articles to cluster different topics within a subject field (Guo et al., 2021; Rosokhata et al., 2021). Others use this tool to learn about co-citation patterns (Small, 1973), or understand the subject research lines (Ibarra-Vargas et al., 2023). In the field of bibliometric analysis of retracted articles, the same logic can be applied to understand the characteristics of these publications. While no article was found using any specific algorithm, some studies use the concept of cluster with a similar purpose. Using network data of co-existing keywords, Eldakar & Shehata (2023) studied the main topics in which there are retracted articles in the technology fields. Zhang et al. (2019) created a metric “retracted index” measuring the prevalence of retractions across scientific disciplines, in which clustering was useful to group different fields.

In the analysis of time-to-retraction, a tool that has been extensively used is survival analysis and, more specifically, the Kaplan-Meier model. In an analysis focusing on a sole author, a Kaplan-Meier survival curve was constructed for the time to retraction of the articles published by the author (Saikia & Thakuria, 2019). In another article about three of the authors with the highest number of retractions, the same procedure was applied for each of the authors and the curves were compared with a modified Gehan-Wilcoxon test (Miranda & Garcia-Carpintero, 2019). Other studies do not specify the type of survival model employed. For example, Zheng & Fu (2022) mentions the use of survival models to compare the time to retraction of open access papers compared to limited or non-open access papers. Chen et al. (2013) focused on highly cited retracted articles and applied survival models, measuring the time to retraction, post-retraction citation and citation since publication.

3. MATERIALS AND METHODS

This section delineates the methodologies undertaken by this Master Thesis. To ensure reproducibility, the data sources and pre-processing will be explained, coupled with the research methods employed to obtain the results.

3.1. DATA SOURCES, TABLES AND PRE-PROCESSING

This study draws data from three primary sources. The inclusion of these sources was guided by previous literature, while still addressing the current limitations of the scope of this analysis.

Retraction Watch Database (RWD): it is the most comprehensive source of information describing retractions in detail. It contains data the characteristics of the retraction, including the reasons, dates, Digital Object Identifier (DOI), if it is paywalled as well as some characteristics of the retracted paper like the authors, country, journal or institution. For its completeness and availability, it is one of the most widely used data sources in this area of study (Fang et al., 2012; Serghiou et al., 2021; Vuong et al., 2020; Zuckerman, 2020).

Scimago (SCI): it contains metrics designed to assess the performance of journals, institutions, media or research centres. From its data, it generates the *Scimago* Journal Rank (SJR) of each journal, adjusting the significance of each citation relative to the importance of the journal. This approach ensures that citations from more influential journals carry greater weight than those from less significant ones. This source was important to quantify how different ranking journals behaved.

Web of Science (WoS): one of the most popular research platforms, which provides access to bibliometric data. It offers data on scientific documents characteristics, including authors, institutions, cited references, number of citations, among others. It was especially important in the characterization of the articles under analysis and to obtain the relevant citation and correction information. The provided data was subsequently enriched with additional information and standard formatting by using the package *bibliometrix* in R (Aria & Cuccurullo, 2017).

The database comprises of three key tables, each of which captures a specific aspect of the data. The **Retractions table** encompasses all relevant information needed to characterize the retracted articles. Three sources of data were joined to create this table: firstly, the RWD was obtained through communication with the source. Then, the retractions were filtered by document type, to include only “Research Article”, “Clinical Study”, “Commentary/Editorial”, “Review Article”, “Meta-Analysis”, “Case Report”, and “Retracted Article”. The article characteristics were collected by searching in WoS for all DOIs pertaining to retractions of articles present in the RWD. This information was merged to the RWD by the DOI of the article.

Lastly, journal data from the year 2022 from SCI was merged to the retractions table by ISSN, to include information on journal ranking and quartiles.

One **Citations table** was created with the information about articles who cited publications from the Retractions table. This information was vital to the analysis of the pre- and post-retraction citations. From the DOI of the articles in the Retractions, the WoS Citation Search was used to obtain the relevant information for each retracted article. Since each observation of this table correspond to a citation to an article, the articles that did not receive any citation are omitted from analysis. Furthermore, citations corresponding to the retraction notice of the article and those whose publication date preceded the date of publication of the retracted article were removed.

Finally, the **Notices table** was created by querying in the WoS platform for the document types “(Correction OR Correction, Addition OR Expression of Concern OR Item Withdrawal OR Retraction)”. It includes all document types that indicate that something needs to be corrected (even if the correction is the removal from the scientific literature). To ensure that all retractions were included in this table, the records of the Retractions table missing in the Correction table were concatenated into the dataset. Then, the data was filtered by matching the referenced papers DOIs to the Citations table. This meant that each correction was related to a citation in the database. To limit the corrections to those that occurred after the retraction of the original paper, the observations were matched by DOI and filtered by Date, so that only corrections published after the retraction are included. The final result was a table with all notices associated to the pre-retraction citations, which were published after the retraction.

The key variable categorizing the articles as either within the biomedicine area or not was done in accordance to the list of scientific areas available in the WoS documentation (*Research Areas*, 2023), in which the areas were manually categorised. The occurrence of misconduct was also manually classified according to the reasons included in the RWD. An approach of plausible deniability was taken, in which cases that were not clearly a sign of rule breaking, the reason would be classified as not being misconduct. Additionally, since the RWD included a large number of reasons, sometimes paraphrases of one another, a variable was created grouping reasons into categories. Both the classification of the biomedical areas and the reasons within misconduct are detailed in the Appendix A.

An additional pre-processing step was in dealing with missing values. In WoS data, the date of publication is divided into two separate variables: year of publication and date (which included day and month). For those observations in which date was missing, the date was filled with the 1st of January. The actual date of the event was considered to be less relevant to the analysis than the year, so it would not add much noise to the analysis. Fortunately, the year of publication had no missing values.

Diagram 3-1 represents the preprocessing steps taken.

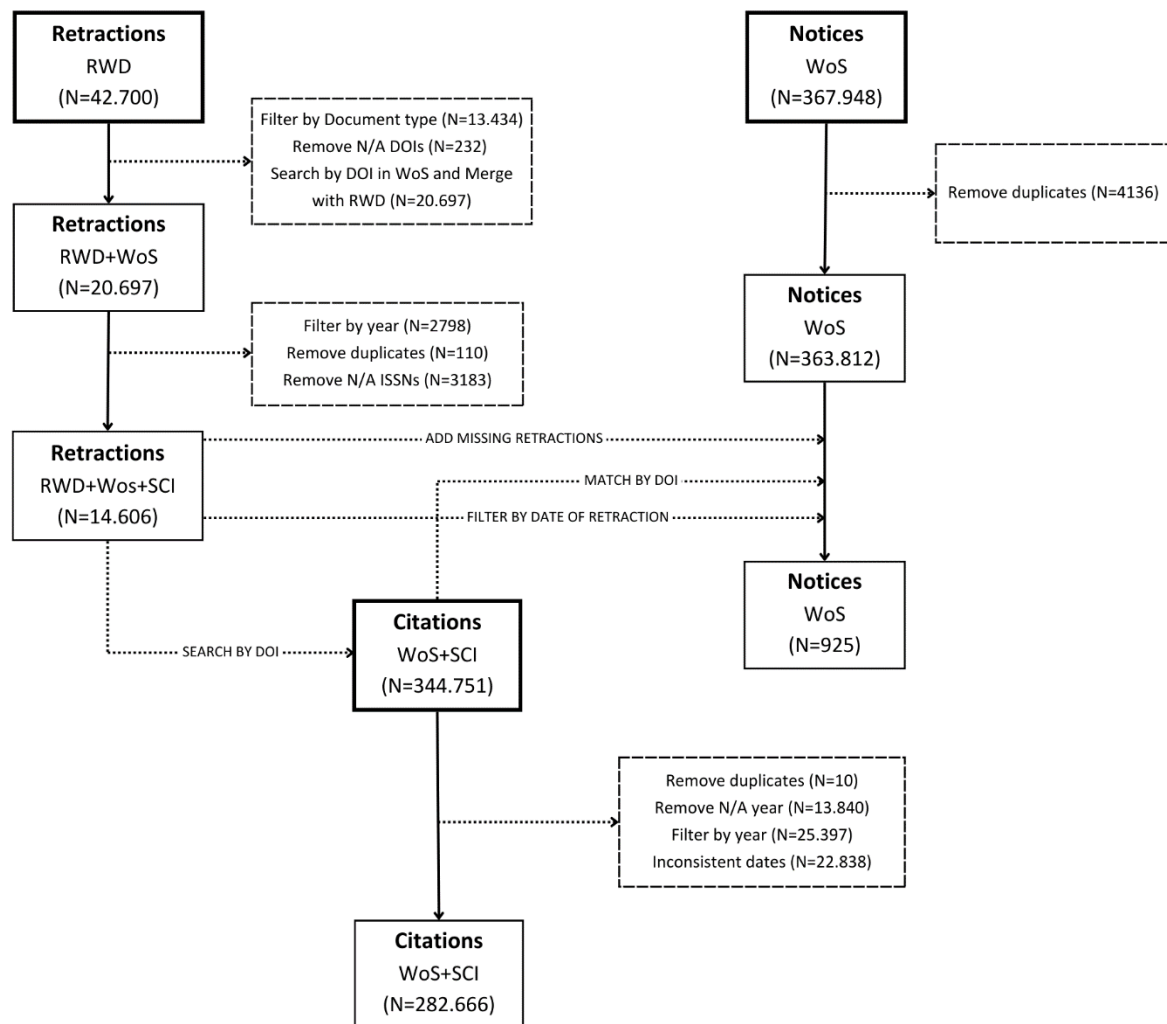


Figure 3-1. Methodological approach¹

In total, three tables were used in the analysis. The retractions dataset comprised of 14.606 observations, the citations dataset included 282.666 observations, and the corrections dataset included 925 observations. Table 3-1 presents descriptive statistics regarding the variables of importance.

¹ Bold outline of the tables indicates the starting point of the data.

Table 3-1. Descriptive statistics numerical key variables

Variables	Avg.	Min.	Median	Max.	Std. Dev.
Retraction Date	31/01/2018	01/03/2000	29/08/2019	31/12/2022	
Original Paper Date	29/01/2015	01/02/1993	10/05/2016	17/11/2022	
Total times cited	23	1	9	3982	61
Time to Retraction (TTR)	36	0	23	258	39
Number of retractions by first author	5	1	1	113	14

Table 3-2 represents the descriptive statistics of categorical variables.

Table 3-2. Descriptive statistics categorical key variables

Variable	Unique	Top	Frequency
Country	143	China	4932
Author	52899	Hironobu Ueshima	113
Reason	18	Unreliable	4804
Quartile	4	Q1	8707
Scientific Area	6	Biomedicine	6029
Misconduct	2	Yes	11990
Biomedicine	2	No	8577

The last two decades have seen an exponential growth in the number of retractions ($R^2 = 0.95$; $p\text{-value} = 6.4 \times 10^{-13}$), shown in figure 3-2. Starting at the turn of the century, in the year 2000 only 13 articles were retracted. In 2010, that number had skyrocketed to 319, followed by almost a twofold increase in 2015 (745) and, in our last year of analysis, 2022, an astounding 3255 retractions were issued.

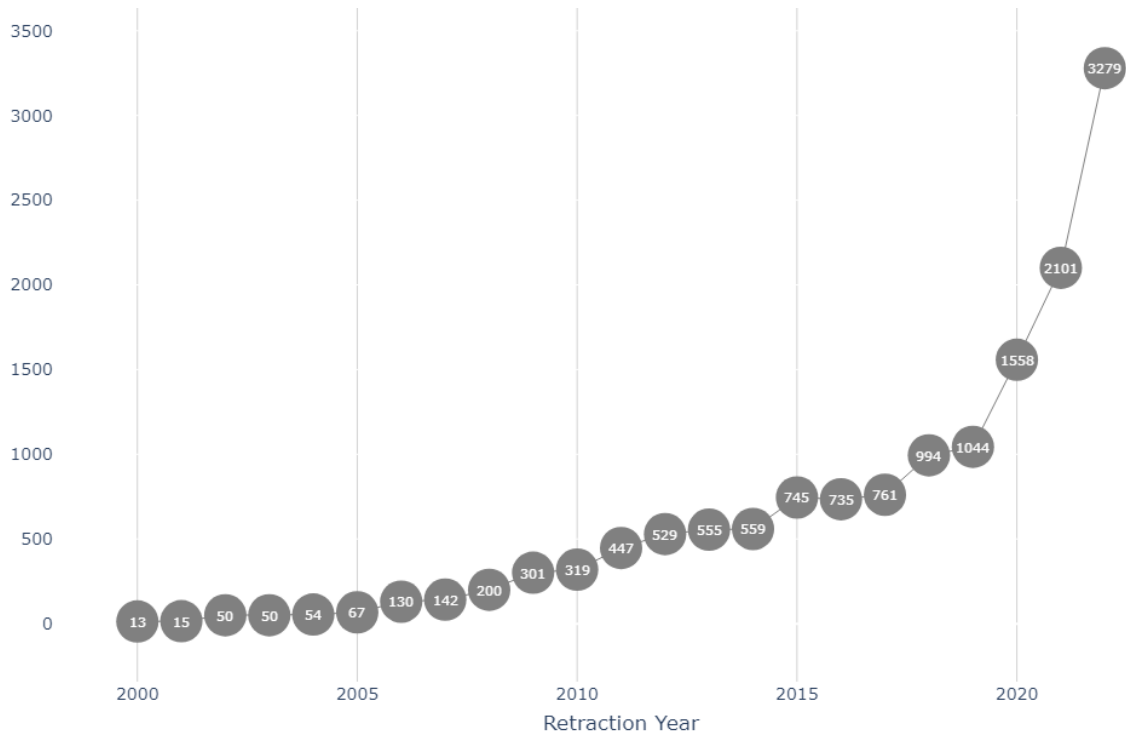


Figure 3-2. Evolution of number of article retractions

3.2. ANALYTICAL METHODS

A top-down approach was used to achieve the objectives. Both use descriptive statistics, statistical testing and visualisations as a means to understand disparities between scientific areas. Further tools were applied to each subsection, which would strengthen the understanding of the relationships.

To study the time to retraction of the papers, the time between the publication and retraction were calculated. After the basic foundations of statistical and visualisation evaluation, it was opted to formalize the analysis of the time to retraction through a Kaplan-Meier survival model. To measure the magnitude of impact of the scientific area and misconduct, a Cox-Hazard Proportional model was implemented. The Cox regression is formalised as follows:

$$(1) H(t) = H_0(t) \times \exp[\beta_1 \text{misconduct}_1 + \beta_2 \text{biomedicine}_2] + u$$

Both the Kaplan-Meier and the Cox Proportional Hazard model are types of survival models, which capture the time until a certain event occurs. It can be applied to the time-to-retraction, to obtain a measure of “survival” of the articles from the moment of their publishing up to their retraction (Goel et al., 2010; Zheng & Fu, 2022). Annex A explains the theoretical background on this topic.

The post-retraction citations analysis included a Negative Binomial regression for panel data Greene (2003), with the objective of understanding the relationship between the yearly

number of citations and the year of retraction. This model was deemed the best fit, as it accounts for the count nature of the target variable and the overdispersion present in the data, which if unaccounted for, would lead to inaccurate and inefficient results (Ghazal & Zulkhibri, 2015). The theoretical background on this topic is available in the Annex B.

The number of citations an article received at a given time t was modelled against several variables. The time difference t , in years, from the publication of the citation to the retraction year was used to measure the evolution of the citations both in their natural progression and to measure the effect of the retraction on the citations. The variable takes negative values until the year of retraction, which is encoded as zero, and from then on is positive. The *misconduct* encodes the occurrence of malpractice as reason for retraction. This binary variable is essential to distinguish the effectiveness of more problematic cases compared to honest errors. Finally, the variable *biomed* was included as the main objective of this thesis is to understand whether the biomedicine scientific area has a different behaviour than other areas. The formalisation of the base model is illustrated below.

$$(2) \log(citations_{it}) = \beta_0 + \beta_1 t + \beta_2 misconduct_{it} + \beta_3 biomed_{it} + u_{it}$$

Following the study of the specification of the model (RESET test p-value < 0.001), several changes were made. Firstly, the time t was included in its quadratic form to account for the diminishing returns of the variable and especially to account for the reduction in number of citations due to the retraction. Two interactions were also included in the model. The *misconduct* was interacted with t , as the reason for retraction could invalidate the findings and lead to scientists avoiding referencing the article. *misconduct* was also interacted with *biomed*, to understand whether the scientific area groups had differing citation behaviour, given the occurrence of misconduct. Thus, the final form specification of the model became:

$$\begin{aligned} (3) \log(citations_{it}) &= \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 misconduct_{it} + \beta_4 biomed_{it} \\ &+ \beta_5 (t \times misconduct_{it}) + \beta_6 (t \times biomed_{it}) \\ &+ \beta_7 (biomed_{it} \times misconduct_{it}) + u_{it} \end{aligned}$$

It is worth noting that in this analysis only retracted articles with indexed citations in the WoS database were included. In other words, for each retracted article represented, everyone included at least one non-zero citation count within the considered time span. Additionally, the first time t for any given article is the date of publication of the retracted article.

For the analysis of the corrective action of pre-retraction citations, the analysis was separated by the three main document types: corrections, retractions and expressions of concern. There were a small number of withdrawals which were considered as retractions within the context of this analysis, since they share the same consequences, and it would not make sense to isolate so few cases. For the retractions and expressions of concern, the number of instances were compared for different subsets of the data.

In the analysis of corrections, a sample was chosen for individual analysis. For that purpose, a stratified proportional sampling method was used, for the estimation of proportion of corrections that mentioned the retracted article as reason for correction². This sampling design is especially adequate when one needs a representative sample for sub-groups of the population, in this case, the biomedical area. So, a random sample of corrections was selected in the biomedical and non-biomedical field. To facilitate the search, only corrections with available DOIs were eligible to be selected. Then, each of them was categorised as either mentioning a retracted citing reference or not.

While all the previous information is useful on its own, by joining the information from the first and second objective, it is possible to draw “profiles” of the retracted articles, and understand whether retracted articles displayed different characteristics depending on the scientific area. Since the chosen variables had a mixture of data types, the approach proposed by van de Velden et al. (2019) was undertaken. It encompassed the implementation of the K-Medoids algorithm with the Gower distance to measure similarity between observations. To choose the best number of clusters, a combination of the elbow method and empirical knowledge of the problem were used. A detailed explanation of the methods is available in Annex D.

The set of characteristics considered reflected several facets of concern: number of authors, if the article was retracted for misconduct, the number of retractions of first author (discretized), the average number of citations per year after retraction, and time to retraction. The number of authors and number of affiliations were included in the analysis, as it has been studied that authors whose work was retracted for unethical behaviour are more prone to collaborative work. Furthermore, the number of retractions of the first author were used as a measure of the problem being the fruit of negligence or a deeper problem. The first author was chosen, as it is usually the one offering the biggest contribution to the study. The binary variable indicating the occurrence of misconduct was included to capture the impact of such serious cases. Finally, the two last variables pertain to the second objective, previously discussed as the way of measuring concern.

² Please consult Annex C for more information.

4. RESULTS AND DISCUSSION

4.1. RETRACTION CHARACTERISTICS

In total, biomedicine accumulated 6029 retractions, whereas all other areas amassed 8577 instances. It is possible to further decompose the scientific areas, which shows that biomedicine is leading in number of retractions, with more than a thousand retractions gap to the technology, the second most retracted area. Most retractions included a single scientific area, with the exception of the combination between biomedicine and life sciences, with 1122 retractions, and physical sciences and technology with 724 retractions. Figure 4-1 shows the decomposition of retractions per area.

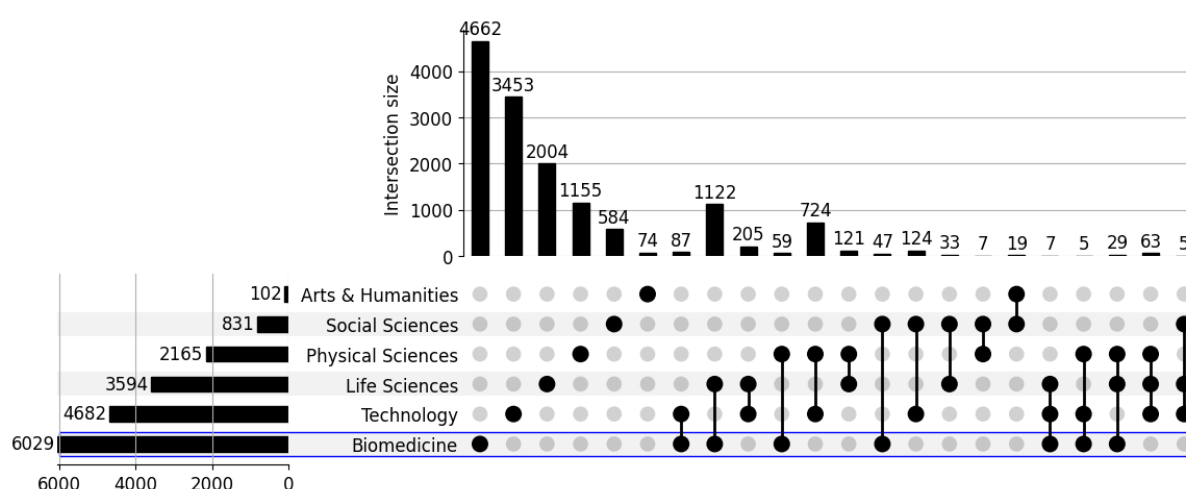


Figure 4-1. Scientific Area combinations³

Every scientific area with the exception of the Arts & Humanities, showed a growing the number of retractions (see figure 4-2). The rise of the number of retractions was led by the biomedical field, with a steady increase 2020, followed by an abrupt spike. One possible reason for this is the increased number of publications regarding the COVID-19 pandemic. With the higher incentive to publish significant scientific discoveries both from the perspective of authors and editors lead to an increased rate of retractions (Schonhaut et al., 2023). The technological and life sciences areas had an evolution more akin an exponential growth. This may reflect that the governance of this self-correction mechanism has expanded earlier in some areas than others. This hypothesis is supported by the growing number of journals that have issued retractions and the stable number of retractions per journal each year.

³ Only combinations of scientific areas that appear five times or more are represented in the vertical bar chart. The combinations are encoded by the connected shaded circles. The horizontal bar chart includes every retracted article associated with each scientific area.

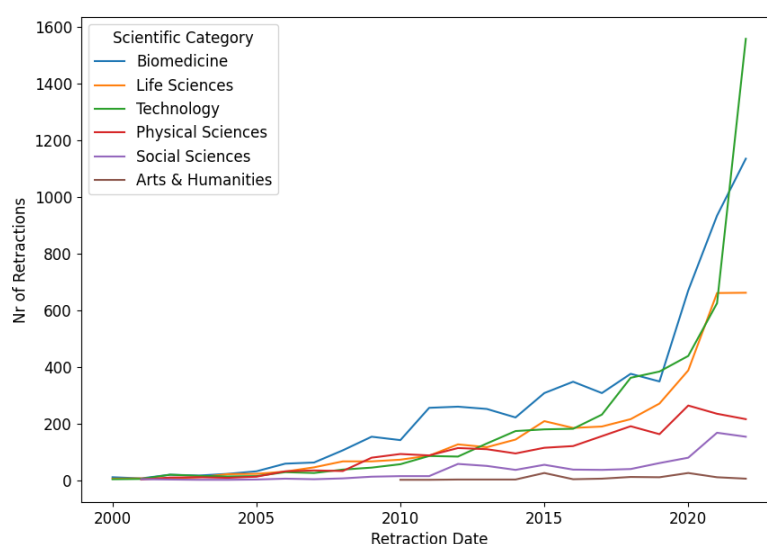


Figure 4-2. Evolution of number of retractions and journals according to scientific area

Considering the grouping of other areas other than biomedicine, there has been a rise in retractions over the years in both groups of scientific areas, shown in figure 4-3. And while the absolute number of retractions has been consistently lower in biomedicine, the rate of retractions has always been higher. Between 2015 and 2019, the gap between these two groups started to shorten, as the rate of biomedical retractions started to stabilise. In 2020 the number of retractions skyrocketed. In other areas, the rate of retractions increased more gradually until the last year of analysis, where it almost doubled. The practice of retracting articles has been more widespread in the field of biomedicine compared to other disciplines. This trend is evidenced by the fact that the yearly number of journals issuing retractions the biomedical field matched the number of journals of all other areas combined until 2013⁴. However, this discrepancy does not indicate a higher incidence of flawed research in the biomedical field, as the rate of retractions per journal was comparable across both groups.

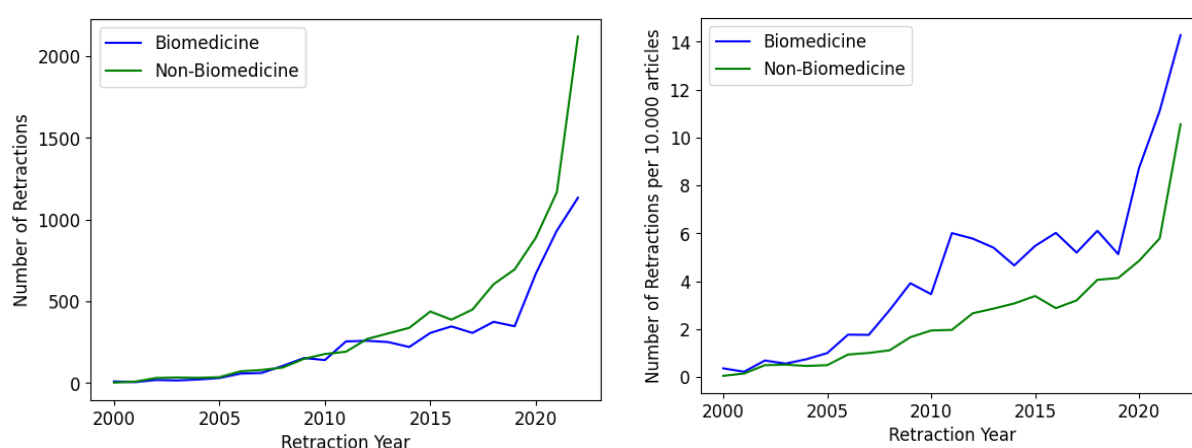


Figure 4-3. Rise in number of retractions per year

⁴ This information is illustrated in Appendix B.

Among the most common reasons were the retraction for unreliable results, with 4804 cases, followed by retractions as a result of investigations, with 4767 cases, and partial or total duplication with 4321 cases (figure 4-4). The top three reasons presented identical order and normalised frequency for both biomedicine and other areas. Naturally, in the biomedical area, the number of cases related to patient concerns were higher. Despite non biomedical areas having more retractions overall, the absolute number of articles retracted for fabrication/manipulation, general misconduct, and paper mills recorded a significantly higher number of instances. The most evident gap is in the number of retractions regarding issues with the peer review, which constitutes 22% of the total cases in the non-biomedical areas, but only 4% in the biomedical field⁵.

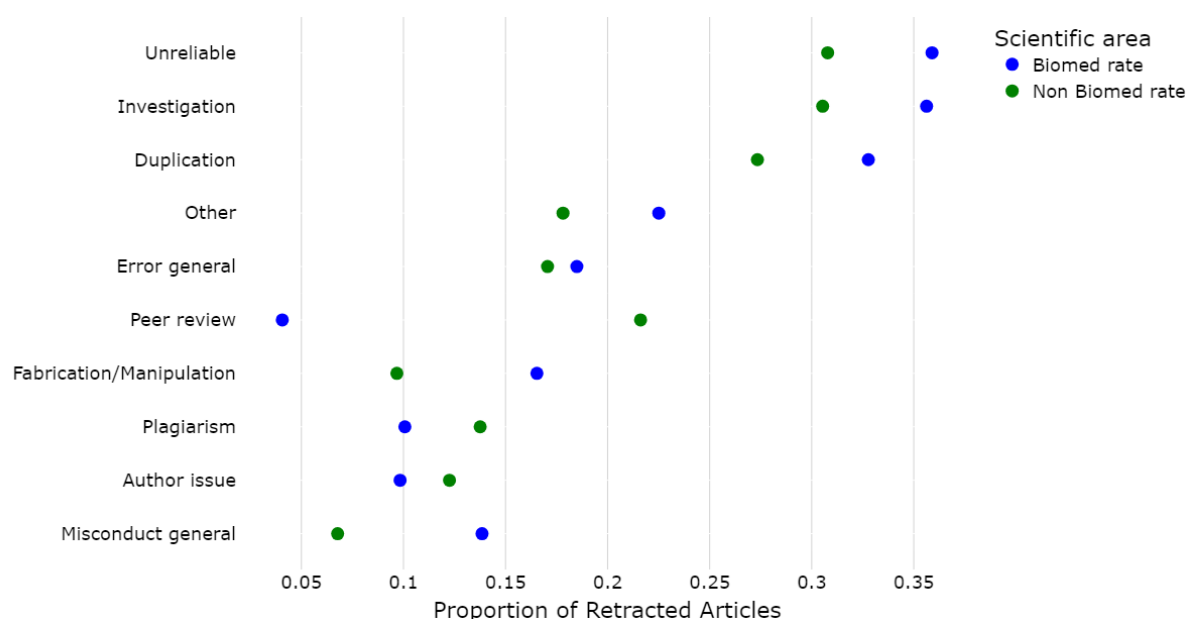


Figure 4-4. Reasons for retraction as a proportion of total retraction articles

Additionally, articles retracted for misconduct have a higher incidence than articles retracted due to other reasons, with 11990 instances from the former compared to 2616 from the latter. In the biomedical area, 81% of retractions were associated with misconduct, compared to 83% in the non-biomedical areas. Considering the date of retraction, both subgroups increased steadily until 2019, showing a start of a tendency to converge. The pattern was broken in 2020, as the number of retractions for misconduct escalated (see figure 4-5). This pattern was found for both scientific areas. In retractions without indication of misconduct, the non-biomedical fields saw a slight decrease in retractions, while in the biomedical field it remained almost constant.

⁵ Please consult the Appendix B for the complete list of reasons and their share. Note that the percentage may add up to more than one, as retraction notices may list multiple reasons.

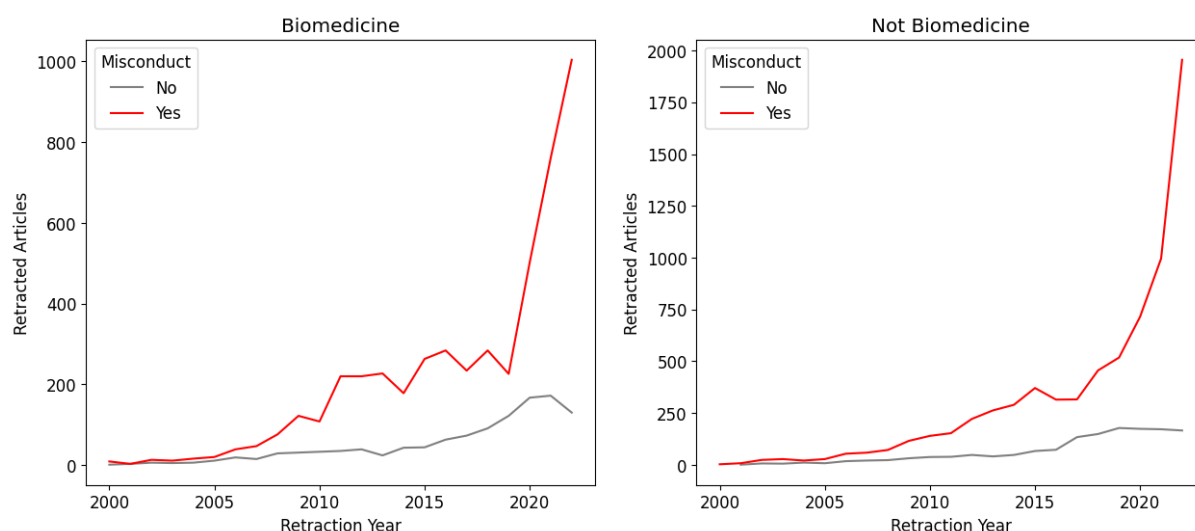


Figure 4-5. Evolution of misconduct

Within the authors that most published retracted articles between 2000 and 2022, the vast majority of articles were retracted for misconduct. From the top 15 most retracted authors, only four did not publish any article in the biomedical area, and from the remaining ten, only one did not publish the majority of their work in this area. Additionally, four out of the five top retracted authors were anaesthesiologists. The third place is reserved to Ali Nazari, studying structural materials. Hironobu Ueshima accumulated a total of 113 retractions, all of them due to misconduct, making him the most prevalent retracted author.

Another facet analysed was geography, illustrated in figure 4-6. The vast majority of articles were from authors affiliated to a single country, and the one with most retractions was China, followed by the United States. It is also worth noting that these countries the highest frequency article producers (NSF, 2023). It is expected that the number of retractions from China will increase significantly in 2024 as a result of the nationwide audit for research misconduct, following the mass retractions from the publisher Hindawi (Mallapaty, 2024). Comparing groups of countries, it was concluded that BRICS accumulated the highest number of retractions, followed by the countries that compose G7, and then all other countries combined. However, for most of the time range considered, the number of retractions from G7 countries was higher than BRICS countries for both groups of scientific areas. However, in 2020 for biomedicine, and 2018 for other areas, the number of retractions from BRICS countries shot up exponentially, surpassing those of G7. The 2020 case can be explained by the pandemic and the race to find an effective vaccine. For the 2018 case, a possible explanation was not found.

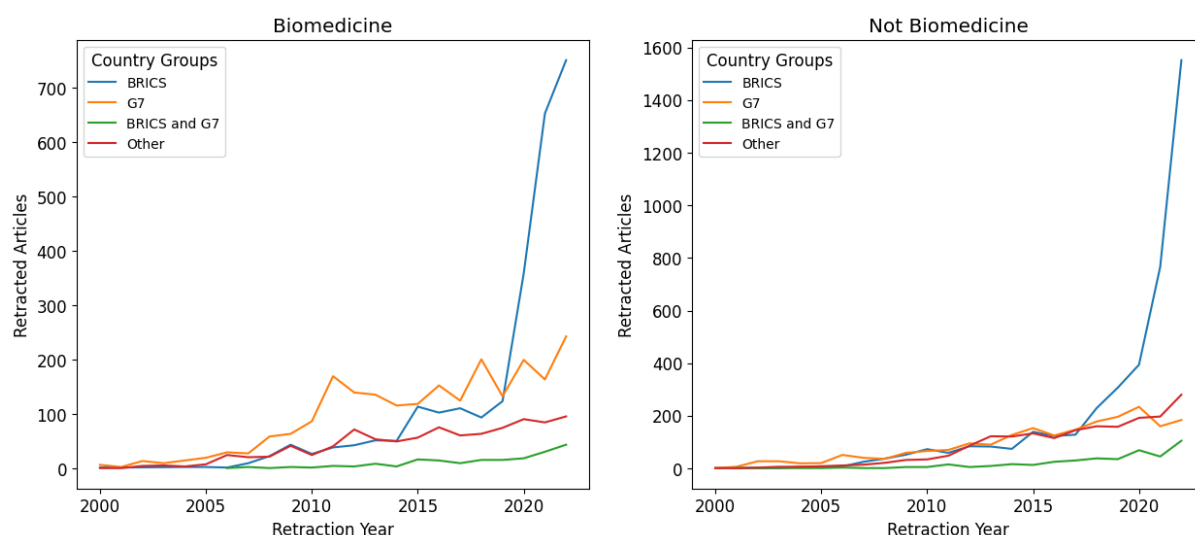


Figure 4-6. Retractions in groups of countries

4.2. TIME-TO-RETRACTION

The average time an article was in press before retraction was 36 months, equivalent to three years. This contrasts to previous literature, (Budd et al., 1998; Redman et al., 2008), by almost a year for the most recent analysis. The biomedical area had a significant higher time to retraction compared to other areas ($p\text{-value} < 0.01$), although it is worth noting that not only is the standard deviation of the biomedical area higher, but also that both have a much lower median time-to-retraction (TTR). In fact, this skew in the distribution distorts the view of the centrality measure by more than a year. Table 4-1 describes the centrality and dispersion measures associated with the TTR.

Table 4-1. Time to retraction descriptive statistics

Time to Retraction	Obs.	Avg.	Median	Min.	Max.	Std. Dev.
Total	14606	36	23	0	258	39
Biomedicine	6029	41	26	0	258	42
Non Biomedicine	8577	33	22	0	256	36

Overall, there is a clear pattern of reasons associated with misconduct (AVG = 39 months, MED = 26 months), taking more time than others (AVG = 22 months, MED = 11 months). In both scientific areas, the fabrication/manipulation of data was the slowest to retract. Other forms of misconduct like plagiarism among the fastest to retract on average, third only to salami slicing and peer review issues. Table 4-2 breaks down the average time to retraction for each reason.

Table 4-2. Average time to retraction (in months) by reason

Reason	TTR Total	TTR Biomed	TTR Non Biomed
Fabrication/Manipulation	63	67	57
Misconduct general	57	66	44
Patient concerns	52	55	42
Duplication	48	48	47
Investigation	45	55	37
Unreliable	41	42	39
Miscommunication	36	41	31
Other	34	39	30
Paper mill	34	35	31
Author issue	30	40	25
Referencing	30	51	25
Contamination	28	29	28
Error general	28	28	27
Conflict third party/institution	27	32	24
Legal	27	32	23
Plagiarism	27	28	26
Peer review	25	17	26
Salami Slicing	20	18	21

The Kaplan-Meier model provided additional information. At the one-year mark (12 months), 32% of articles subject to retraction will have been removed. At 24 months, that number increases to 53%, and at 36 months two thirds of articles will have been retracted. From the survival function is possible to understand that there is no singular point at which retractions are more likely to happen, it is a gradual process with no break in the survival distribution. Figure 4-7 shows the Kaplan-Meier survival curve.

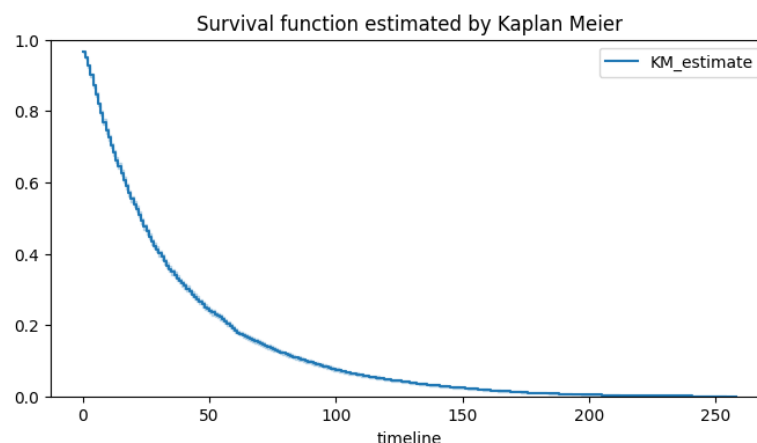


Figure 4-7. Kaplan-Meier survival curve of retracted articles

To understand the magnitude difference between biomedical and other areas, a Cox Hazard model was implemented. Results indicate a reduced hazard risk in the biomedical area, i.e., that it is slower to retract. Similarly, misconduct is associated with higher time to retraction, with a stronger effect than the scientific area. Additionally, articles in biomedicine take, on average, 22% longer to retract than other areas. Similarly, retractions caused by misconduct are 1.79 times slower to retract. The hazard ratio (HR) can also be interpreted as articles in biomedicine are 18% more likely to not be retracted, and articles without misconduct are 56% more likely to be retracted. The differences between categories in both variables are statistically significant, shown in table 4-3.

Table 4-3. Cox-Hazard Ratio Model Summary⁶

	Biomedicine	Misconduct
Coefficients (COEF)	-0.20	-0.58
Hazard ratios (HR)	0.82	0.56
Std. Error	0.02	0.02
COEF CI	[-0.24, -0.17]	[-0.63, -0.54]
HR CI	[0.79, 0.84]	[0.53, 0.58]
z-score	-12.12	-26.86
p-value	<0.005***	<0.005***
n = 14606	Partial AIC 250147.78	Log-Like Ratio test 775.00 on 2 dof

Analysing the temporal tendencies of the time to retraction, it was noticed that there was a gradual decline from the year of the publishing of the article to the retraction (figure 4-8). This suggests that scientists are being more prompt in their action towards recognising the need and taking action. It may also be an indicator of the reduction of the stigma around retractions, as something that is not necessarily negative. Additionally, the biomedical and non-biomedical fields portray an identical behaviour in regard to the evolution of time to retraction.

⁶ Note that “*”: p-value < 0.10; “**”: p-value < 0.05; “***”: p-value < 0.01.

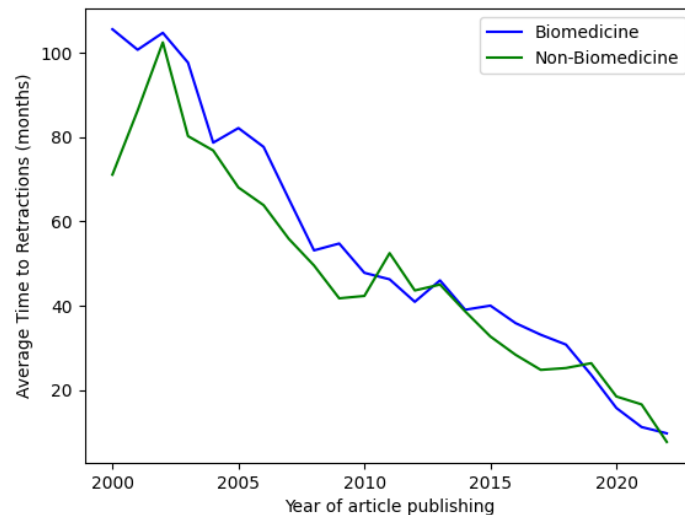


Figure 4-8. Evolution of time to retraction by publishing year

At the same time, and perhaps counterintuitively, more recent retractions have a higher time difference between publication and retraction (see figure 4-9). It is understandable as follows: for any given year, the time from the publishing of an article until the moment it is perceived to be invalid has been decreasing. On the other hand, retractions that are being published more recently have a higher time to retraction, which shows that scientists are reaching further back to correct their mistakes. This is a great indication that the increase in number of retractions in recent years is not necessarily a reason for the concern of a rise in false science, but instead is an indicator of the efficacy of the monitoring systems set in place. Additionally, it corroborates the hypothesis made by Steen et al. (2013) that rising misconduct is partially due to increased scrutiny. Similarly to the previous graph, both groups of scientific areas show a very similar behaviour.

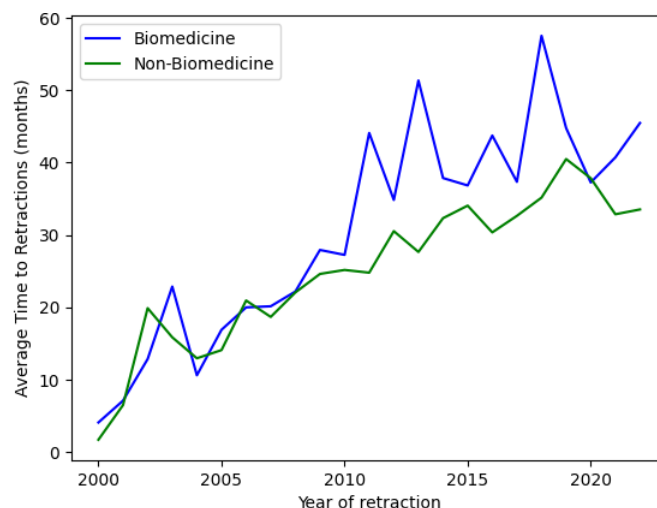


Figure 4-9. Evolution of time to retraction by retraction year

This pattern also is seen in articles retracted for misconduct and other motives. While the average time to retraction has been decreasing for recently published articles, the retractions

being issued have been increasing in gap between date of publishing and retraction. Interestingly, the non-misconduct articles in biomedicine saw a decreased TTR from 2017 to 2019, while for the same years there was an abnormal spike in misconduct articles of the same area.

4.3. POST-RETRACTION CITATIONS

Each article accumulated an average of 23 citations, reducing from 19 citations in the pre-retraction period to 11 citations post-retraction. In the biomedical area, the average number of citations accumulated by an article is 25, and in other areas it is 22. Among the most cited articles, six out of ten were included in the biomedical area. Considering solely the citations post-retraction, the top ten articles who accumulated the most citations were half from biomedicine and other areas. Within this subset, only three articles did not have a decline in their number of yearly citations after retraction (red line in figure 4-10): the “Primary Prevention of Cardiovascular Disease with a Mediterranean Diet”, "Brain hydrogen sulfide is severely decreased in Alzheimer's disease" and “Carbon Nanotubes Are Able To Penetrate Plant Seed Coat and Dramatically Affect Seed Germination and Plant Growth”.

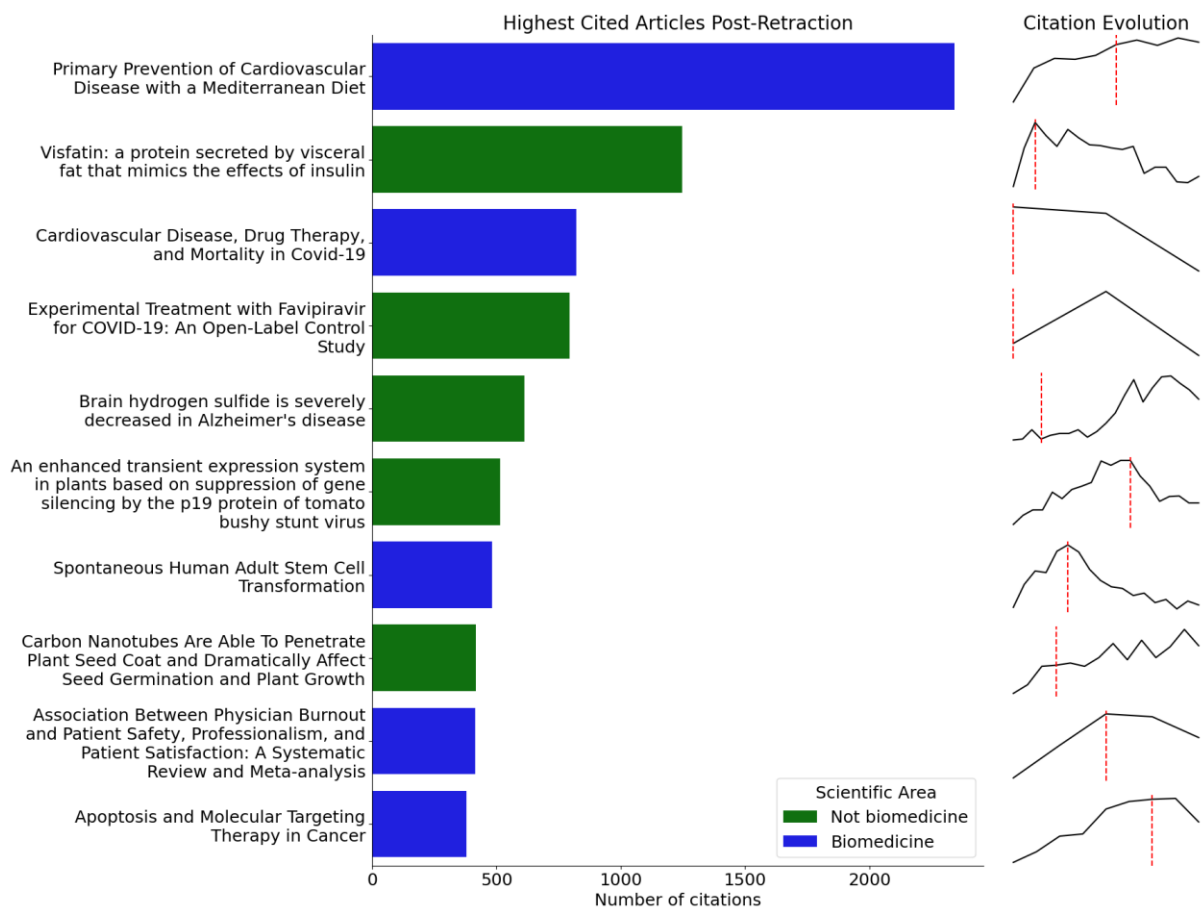


Figure 4-10. Top cited articles post-retraction⁷

It can be seen that the number of citations an article receives increases faster in areas other than biomedicine, but at the time of retraction, both are on par. Then, there is a fast and significant drop in citations. While the average number of citations in the year previous to retraction is four, it drops to one on the year after retraction, and continues to decrease further. Taking the average of the variation in citations per year, the variation is positive before retraction and then negative after. The first year following retraction is the one with the most significant effect. This pattern appears in both groups of scientific areas, similarly. In Figure 4-11, $t = 0$ represents the year of retraction. The graph on the left shows the number of yearly citations an article receives, in relation to the time of retraction. On the right, the yearly variation of citations is shown in relation to time of retraction.

⁷ A table with the full citations corresponding to each article is available in the Appendix B

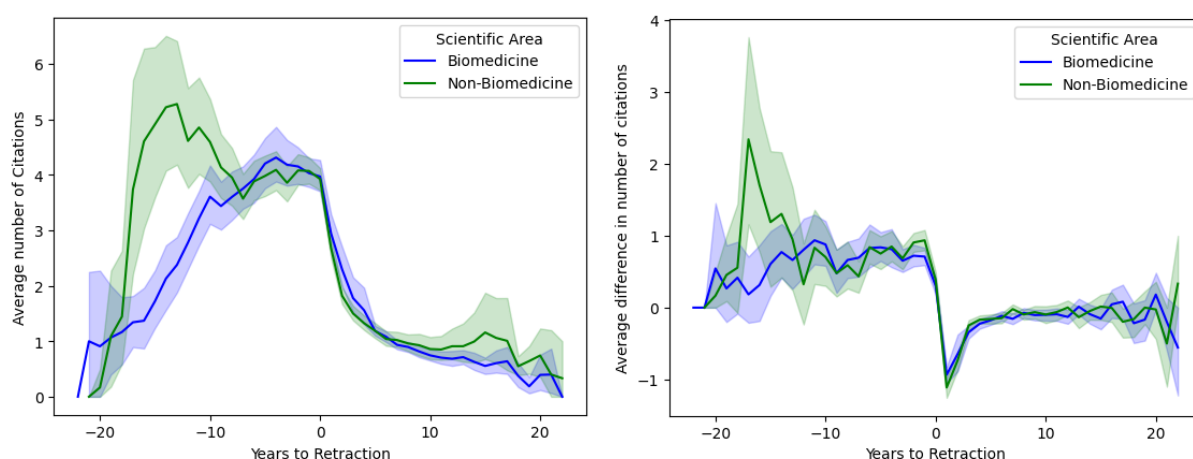


Figure 4-11. Evolution of citations in relation to retraction year

Replicating the analysis for the subset of articles retracted for misconduct, it can be seen that the citations of misconduct articles are higher than others in the pre-retraction period but have a steeper drop at the moment of retraction (figure 4-12). While both scientific areas show the same tendency, the citations of misconduct articles are higher in the non-biomedical articles.

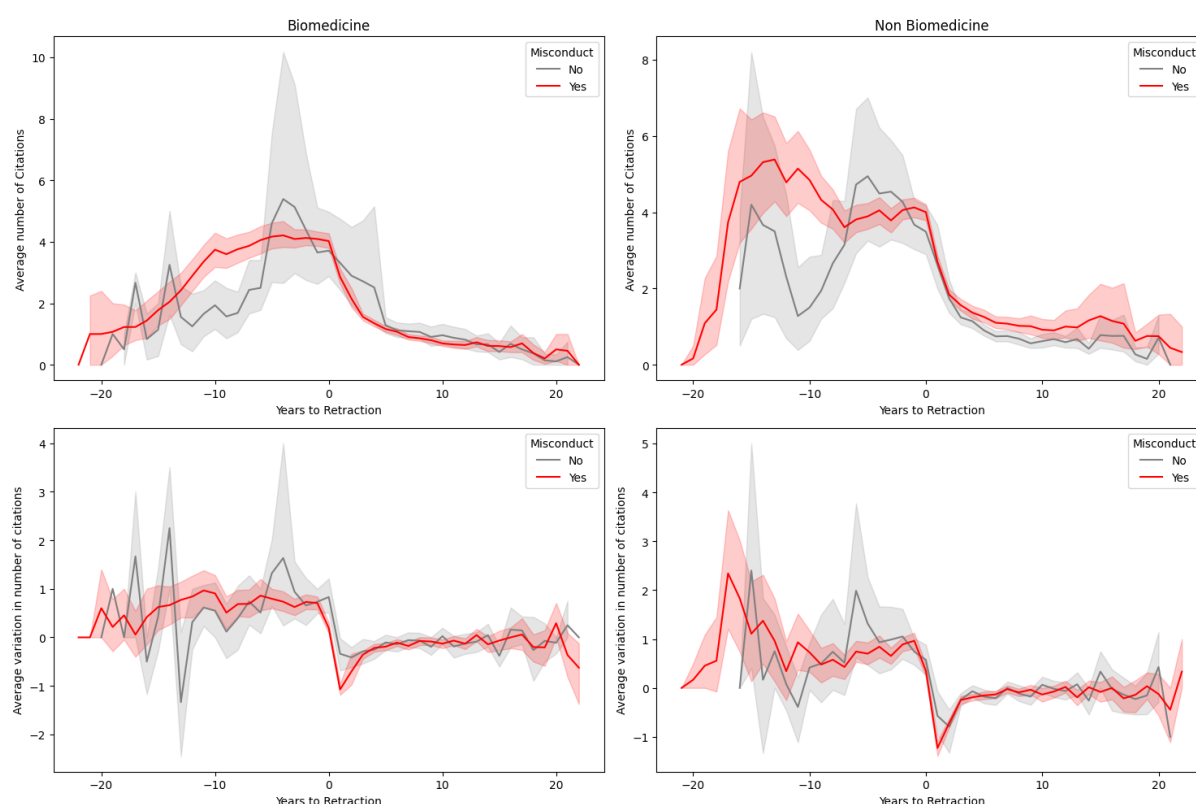


Figure 4-12. Misconduct vs. Biomedicine number of citations

These results are formalised through a regression on the number of citations per year for each retracted article, summarized in table 4-4. Firstly, the baseline average number of citations per year is 2.6 (equivalent to an expected log count of citations of 0.95), which corresponds

to the citations on the year of retraction, from a non-biomedical area and reason not associated to misconduct. Articles citing in the year of retraction might have already been submitted and accepted before the retraction took place, so $t = 0$ it is considered the last year before the fall of citations. The regression shows that citations reach their peak well before the date of retraction, at around 14 years before the retraction.

The interaction of misconduct with time are not statistically significant, indicating that the effect of the time period does not alter significantly depending on the reason for retraction and the scientific area. On the other hand, biomedicine shows a stronger decline in citations, given the factor time. Those two variables are important on their own, both indicating a higher number of citations on average. Additionally, articles retracted for misconduct also have a significantly lower number of citations per year in the biomedical area.

Table 4-4. Yearly citations regression time evolution⁸

Variable	Estimate	Std. Error	p-value
Intercept	0.945	0.040	<0.001***
t	-0.092	0.005	<0.001***
t²	-0.003	0.000	<0.001***
t * Misconduct	0.009	0.005	0.101
t * Biomedicine	-0.007	0.003	0.042**
Misconduct * Biomedicine	-0.306	0.076	<0.001***
Misconduct	0.157	0.041	<0.001***
Biomedicine	0.265	0.074	<0.001***
n = 105203	AIC 420703	Pseudo R² 0.02	Overdispersion 0.49

An auxiliary regression was fitted, replacing the term “t” and its quadratic by a binary variable indicating whether the specified time period was before or after retraction (table 4-5). In this regression, it became even more apparent that the effect of retraction on the drop in citations does not depend on the reason of the citations. This model also showed a statistically insignificant effect of number of citations before and after retraction, taking a given scientific area. Maintaining all other factors constant, the number of yearly citations reduces on average by 0.48 after retraction. The variation increases to a 0.60 decline for biomedicine, 0.57 for misconduct, and 0.55 for both.

⁸ Note that “*”: p-value < 0.10; “**”: p-value < 0.05; “***”: p-value < 0.01.

Table 4-5. Yearly citations regression of comparing pre and post retractions

Variable	Estimate	Std. Error	p-value
Intercept	1.262	0.063	<0.001***
Is retracted	-0.739	0.074	<0.001***
Is retracted * Misconduct	0.023	0.078	0.772
Is retracted * Biomedicine	0.023	0.033	0.499
Misconduct * Biomedicine	-0.266	0.083	0.001***
Misconduct	0.150	0.067	0.024***
Biomedicine	0.220	0.081	0.006***
n = 105203	AIC 425823	Pseudo R² 0.01	Overdispersion 0.45

4.4. CORRECTIONS TO PRE-RETRACTION CITATIONS

There were no instances where the correction mentioned a retracted article which was not the paper itself. There were instances where the correction did not accurately match the article it. In the biomedical area the sample was reduced to 15 corrections, and in other areas it was reduced to 18. There were also several instances where the retracted article was not indexed as such in either WoS or RWD. In the non-biomedical area, 8 out of 24 cases were retractions labelled as corrections in WoS and not present in the RWD with DOI. In the biomedical area, there were 4 out of 23 cases. This is an indication of the flaws in the current repositories of metadata about articles.

4.5. HOLISTIC VARIABLE ASSESSMENT

To understand how the previously examined characteristics fit together, a Cluster Analysis was performed. Given the results of the previous section, the number of corrections to citations was not included in the analysis. It was opted to choose three clusters, guided by the elbow method and the result which provided the best interpretation. Table 4-6 shows the average of the variables in each cluster. For categorical variables, each row represents a category and the proportion of the cluster belonging to said category.

Table 4-6. Clusters average per variable

	Single case misconduct	Multiple case misconduct	Non misconduct
Number of authors	4.76	5.06	4.87
Average post-retraction citations	0.85	1.15	0.95
Time to retraction	30.96	52.20	18.90
Misconduct No	0.00	0.04	1.00
Misconduct Yes	1.00	0.96	0.00
Number of offenses 1	0.95	0.00	0.84
Number of offenses 2-10	0.00	0.82	0.12
Number of offenses >10	0.05	0.18	0.04
Biomedicine	0.38	0.45	0.44

Cluster “Single case misconduct”

Exclusively composed of misconduct articles, it represents those whose first author was subject to a single retraction. These articles had an average of 30.96 months until retraction. Additionally, it is the cluster with the least amount of yearly post-retraction citations.

Cluster “Multiple case misconduct”

The articles composing this cluster have the highest time to retraction, with more than five years on average. These articles are the ones that most receive post retraction citations, and are predominantly a result of misconduct. This group is also composed exclusively of articles whose first author accumulated more than one retraction.

Cluster “Non misconduct”

These articles were all retracted for reasons not associated with misconduct. They have the lowest time to retraction, with an average nearing 19 months. While the majority of authors suffered a single retraction, 16% issued more.

Given that 41% of all retractions are from the biomedical area, the cluster “Single case misconduct” is the only one with a proportionally smaller representation of this area. However, none of the clusters stands out as having a larger share of observations in one specific area. Additionally, there are two types of misconduct articles, in which one has a significantly higher impact than the other. Cluster “Single case misconduct” is faster to retract and from an author that is not a multiple offender. On the other hand, cluster “Multiple case misconduct” represents the retractions that take a long time to retraction from authors with multiple instances. The time may be a consequence of the number of retractions, as one may

investigate associated work, which will take more time. However, the first group has the lowest average yearly post-retraction citations, and the latter has the highest rate.

5. CONCLUSIONS

This thesis aimed to determine whether the biomedical field demonstrates a heightened concern for correcting literature, through retractions, compared to other scientific domains. The two main objectives entail: (i) identifying temporal, geographical, journal trends of papers citing retractions, according to their scientific area; and (ii) quantifying efforts to correct citations of retractions.

Biomedicine has accumulated more retractions than any other area. And while this area had a gradual increase until 2020, the second most prevalent area, technology, had an exponential growth, surpassing all others on the last year of analysis. Since it could be a scale distortion, the rate of retractions per 10.000 articles published was analysed. Biomedicine still leads for every year of analysis. But while other areas had a consistent rise, the rate in biomedicine showed signs of stabilising until the year 2020. It was also found that the growth of retractions results from an expansion of this mechanism to different journals, i.e., making retractions a more standard practice. The biomedical area stands out by the yearly number of journals issuing retractions which, for more than half of the study period, rivalled the journals of all other areas combined. Still, the rate of the articles per journal remained similar for both groups.

Yet not all retractions raise the same level of concern. While the top reason, unreliable results, is not considered misconduct, for the most part, it is followed by retractions resulting from investigations and duplication. In fact, the cases of misconduct represent 82% of overall instances. For the biomedical area, it represents 81% of the total issues and 83% in other areas. These cases are important to be removed from circulation speedily, to avoid contamination of literature and potential harm to patients. The average time to retraction is 36 months, with biomedical articles taking longer on average. Misconduct-related retractions take more time than those for other reasons, although there has been a trend towards quicker retractions, indicating improved responsiveness in the scientific community.

Retracted articles continue to be cited post-retraction, though they drop significantly. Biomedical articles tend to have slightly higher citation rates than others, but its citations drop more speedily. Misconduct-related articles, on the other hand, accumulate more citations but do not experience a different rate of decline in citations after retraction. The analysis of the corrections to the pre-retraction citations did not follow as expected, but still some insights are able to be taken. It was evidenced that the metadata repositories needed constant maintenance and have its flaws which can lead scientists to unknowingly citing articles that may be retracted. These occurrences are equally likely to happen in the biomedical field or not.

What does this mean for the scientific community? Until the year of the pandemic, the incidence of misconduct had stabilised in the biomedical area, at the same time as the systems governing retractions have expanded. Other areas have shown a substantial growth still,

indicating that it is still evolving. And while the majority of retractions are a result of behaviour deviancy, its rise is not necessarily a bad sign. As it was seen, articles are being flagged quicker than ever before along with the concern for reaching further back to remove unreliable works. Additionally, the retractions are mostly effective. While they cannot eradicate post-retraction citations completely, the rate at which they are cited drops to below one per year.

Overall, although the biomedicine's share of the retractions is noteworthy, it is not possible to state that it holds a heightened concern, compared to other areas. The issue is much more nuanced. Some factors indicate a higher concern in biomedicine, like the higher number of journals retracting for the same rate of retractions per journal, or the top post-retraction citation articles being mostly from other areas. But there are also factors in contrast with the statement, including the increased time to retraction, or the fact that the most retracted authors were from biomedicine (a large portion of which were anaesthesiologists). And, especially, there are the factors that are simply not distinguishable or incomparable between scientific areas. For example, the higher rate of plagiarism in non-biomedicine, or higher number of paper mills in biomedicine are both reasons for misconduct, but neither can be considered better or worse. Within this category are the drop in post-retraction citations, which was found to be statistically indifferent for the different scientific areas. Or the analysis of time to retraction that showed both groups are reducing the time they take to retract papers, but have also continued retracting older publications, and fixing the scientific record.

Even though this study aimed to explore retraction data as a whole, there are some limitations to these conclusions. Firstly, data inconsistencies were found in the preprocessing phase that, while taken into account to reduce possible noise, there may still be hidden issues. This regards especially the dates of publication, in which individual analysis of suspicious articles were found to be either wrong or duplicated with conflicting information in the source database. From that investigation, other cases were found in which the citation did not match the article that it said it cited. Additionally, some author names that appeared unformatted in the database, for example, "Åžakir TÃ¼fekÅŸi". These inconsistencies were not possible to be corrected case by case, given the magnitude of the data, and while measures were taken to mitigate them, it is possible that they impacted the analysis. Moreover, the source databases may not have indexed all the retractions, which may also bias results.

Finally, not only is the issue of retractions sensitive topic that may lead to the avoidance of such measures, but cases of misconduct may take a significant more effort to understand that there is a problem. This means that measuring and characterizing retractions is not only a reflection of the publishing actions of scientists, but also the governance and post-publication peer review the scientific community upholds.

BIBLIOGRAPHICAL REFERENCES

- Abd Elhafeez, S., D'Arrigo, G., Leonardis, D., Fusaro, M., Tripepi, G., & Roumeliotis, S. (2021). Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. *Oxidative Medicine and Cellular Longevity*, 2021. <https://doi.org/10.1155/2021/1302811>
- Ajiferuke, I., & Adekannbi, J. O. (2020). Correction and retraction practices in library and information science journals. *Journal of Librarianship and Information Science*, 52(1), 169–183. <https://doi.org/10.1177/0961000618785408>
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/J.JOI.2017.08.007>
- Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy*, 46(9), 1552–1569. <https://doi.org/10.1016/J.RESPOL.2017.07.003>
- Bhatt, B. (2021). A multi-perspective analysis of retractions in life sciences. *Scientometrics*, 126(5), 4039–4054. <https://doi.org/10.1007/S11192-021-03907-0>
- Biagioli, M., Kenney, M., Martin, B. R., & Walsh, J. P. (2019). Academic misconduct, misrepresentation and gaming: A reassessment. *Research Policy*, 48(2), 401–413. <https://doi.org/10.1016/J.RESPOL.2018.10.025>
- Brainard, J., You, J., & Bonazzi, D. (2018). Rethinking Retractions: The largest-ever database of retracted articles suggests the burgeoning numbers reflect better oversight, not a crisis in science. *Science*, 362(6413), 390–393. <https://doi.org/10.1126/SCIENCE.362.6413.390>
- Budd, J. M., Sievert, M. E., & Schultz, T. R. (1998). Phenomena of Retraction: Reasons for Retraction and Citations to the Publications. *JAMA*, 280(3), 296–297. <https://doi.org/10.1001/JAMA.280.3.296>
- Candal-Pedreira, C., Ruano-Ravina, A., Fernández, E., Ramos, J., Campos-Varela, I., & Pérez-Ríos, M. (2020). Does retraction after misconduct have an impact on citations? A pre-post study. *BMJ Global Health*, 5(11), e003719. <https://doi.org/10.1136/BMJGH-2020-003719>
- Charlesworth Author Services. (2022). *Manuscript withdrawal vs article retraction and self-retraction*. <https://www.cwauthors.com/article/manuscript-withdrawal-versus-article-retraction-and-self-retraction>
- Chen, C., Hu, Z., Milbank, J., & Schultz, T. (2013). A visual analytic study of retracted articles in scientific literature. *Journal of the American Society for Information Science and Technology*, 64(2), 234–253. <https://doi.org/10.1002/ASI.22755>

- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer* 2003 89:2, 89(2), 232–238. <https://doi.org/10.1038/sj.bjc.6601118>
- COPE Council. (2019). *COPE Guidelines: Retraction Guidelines*. <https://doi.org/10.24318/cope.2019.1.4>
- Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. *Accounting, Auditing and Finance*, 1(1), 5–8. <https://doi.org/10.23977/ACCAF.2020.010102>
- Dal-Ré, R., & Ayuso, C. (2019). Reasons for and time to retraction of genetics articles published between 1970 and 2018. *Journal of Medical Genetics*, 56(11), 734–740. <https://doi.org/10.1136/JMEDGENET-2019-106137>
- Davis, P. M. (2012). The persistence of error: a study of retracted articles on the Internet and in personal libraries. *Journal of the Medical Library Association : JMLA*, 100(3), 184. <https://doi.org/10.3163/1536-5050.100.3.008>
- Decullier, E., Huot, L., & Maisonneuve, H. (2014). What time-lag for a retraction search on PubMed? *BMC Research Notes*, 7(1). <https://doi.org/10.1186/1756-0500-7-395>
- Dinh, L., Sarol, J., Cheng, Y. Y., Hsiao, T. K., Parulian, N., & Schneider, J. (2019). Systematic examination of pre- and post-retraction citations. *Proceedings of the Association for Information Science and Technology*, 56(1), 390–394. <https://doi.org/10.1002/PRA2.35>
- Eldakar, M. A. M., & Shehata, A. M. K. (2023). A bibliometric study of article retractions in technology fields in developing economies countries. *Scientometrics*, 128(11), 6047–6083. <https://doi.org/10.1007/S11192-023-04823-1>
- Elsevier. (2024a). *Article withdrawal | Elsevier policy*. <https://www.elsevier.com/about/policies-and-standards/article-withdrawal>
- Elsevier. (2024b). *Errata and corrigenda | Editors | Elsevier*. <https://www.elsevier.com/editor/perk/policy-and-best-practice-errata-and-corrigenda>
- Fanelli, D. (2013). Why Growing Retractions Are (Mostly) a Good Sign. *PLOS Medicine*, 10(12), e1001563. <https://doi.org/10.1371/JOURNAL.PMED.1001563>
- 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(6). <https://doi.org/10.1371/JOURNAL.PONE.0127556>
- 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(42), 17028–17033. <https://doi.org/https://doi.org/10.1073/PNAS.1212247109>
- Freilich, J., & Kim, S. (2023). Is the Patent System Sensitive to Incorrect Information? *The Review of Economics and Statistics*, 1–38. https://doi.org/10.1162/REST_A_01353
- Furman, J. L., Jensen, K., & Murray, F. (2012). Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine. *Research Policy*, 41(2), 276–290. <https://doi.org/10.1016/J.RESPOL.2011.11.001>
- George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21(4), 686–694. <https://doi.org/10.1007/S12350-014-9908-2>
- Ghazal, R., & Zulkhibri, M. (2015). Determinants of innovation outputs in developing countries: Evidence from panel data negative binomial approach. *Journal of Economic Studies*, 42(2), 237–260. <https://doi.org/10.1108/JES-01-2013-0016/FULL/PDF>
- Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, 1(4), 274. <https://doi.org/10.4103/0974-7788.76794>
- Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4), 857. <https://doi.org/10.2307/2528823>
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
- Guo, Y. M., Huang, Z. L., Guo, J., Guo, X. R., Li, H., Liu, M. Y., Ezzeddine, S., & Nkeli, M. J. (2021). A bibliometric analysis and visualization of blockchain. *Future Generation Computer Systems*, 116, 316–332. <https://doi.org/10.1016/J.FUTURE.2020.10.023>
- Heen, E., & Vogt, H. (2024). Scientific rot: Unsustainable publishing practices threatens trust in medicine. *Journal of Evaluation in Clinical Practice*. <https://doi.org/10.1111/JEP.13989>
- Heibi, I., & Peroni, S. (2022). A quantitative and qualitative open citation analysis of retracted articles in the humanities. *Quantitative Science Studies*, 3(4), 953–975. https://doi.org/10.1162/QSS_A_00222/2057107/QSS_A_00222.PDF
- Hsiao, T. K., & Schneider, J. (2022). 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), 1144–1169. https://doi.org/10.1162/QSS_A_00155
- Ibarra-Vargas, S. B., Awad, G., & Velasquez, J. D. (2023). A Bibliometric and Cluster Analysis of Blended Learning Literature. *Revista Iberoamericana de Tecnologías Del Aprendizaje*, 18(1), 86–99. <https://doi.org/10.1109/RITA.2023.3250583>

- Ioannidis, J. P. A. (2012). Scientific Credibility: A Fluctuating Trajectory Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), 645–654. <https://doi.org/10.1177/1745691612464056>
- Jin, X., & Han, J. (2011). K-Medoids Clustering. *Encyclopedia of Machine Learning*, 564–565. https://doi.org/10.1007/978-0-387-30164-8_426
- Khademizadeh, S., Danesh, F., Esmaeili, S., Lund, B., & Santos-d'Amorim, K. (2023). Evolution of retracted publications in the medical sciences: Citations analysis, bibliometrics, and altmetrics trends. *Accountability in Research*. <https://doi.org/10.1080/08989621.2023.2223996>
- Kohl, C. B. S., & Faggion, C. M. (2022). A comprehensive overview of studies that assessed article retractions within the biomedical sciences. *Accountability in Research*. <https://doi.org/10.1080/08989621.2022.2154660>
- Kühberger, A., Streit, D., & Scherndl, T. (2022). Self-correction in science: The effect of retraction on the frequency of citations. *PLOS ONE*, 17(12). <https://doi.org/10.1371/JOURNAL.PONE.0277814>
- Kurt, S. (2018). Why do authors publish in predatory journals? *Learned Publishing*, 31(2), 141–147. <https://doi.org/10.1002/LEAP.1150>
- Liu, W., & Lei, L. (2021). Retractions in the Middle East from 1999 to 2018: a bibliometric analysis. *Scientometrics*, 126(6), 4687–4700. <https://doi.org/10.1007/S11192-021-03919-W>
- Malkov, D., Yaqub, O., & Siepel, J. (2023). The spread of retracted research into policy literature. *Quantitative Science Studies*, 4(1), 68–90. https://doi.org/10.1162/QSS_A_00243
- Mallapaty, S. (2024). China conducts first nationwide review of retractions and research misconduct. *Nature*, 626(8000), 700–701. <https://doi.org/10.1038/D41586-024-00397-X>
- Marcovitch, H. (2007). Misconduct by researchers and authors. *Gaceta Sanitaria*, 21(6), 492–499. <https://doi.org/10.1157/13112245>
- Marcus, A., & Oransky, I. (2014). What Studies of Retractions Tell Us. *Journal of Microbiology & Biology Education*, 15(2), 151–154. <https://doi.org/10.1128/JMBE.V15I2.855>
- Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature* 2005 435:7043, 435(7043), 737–738. <https://doi.org/10.1038/435737a>
- Miranda, R., & Garcia-Carpintero, E. (2019). Comparison of the share of documents and citations from different quartile journals in 25 research areas. *Scientometrics*, 121(1), 479–501. <https://doi.org/10.1007/S11192-019-03210-Z>

- Nath, S. B., Marcus, S. C., & Druss, B. G. (2006). Retractions in the research literature: misconduct or mistakes? *Medical Journal of Australia*, 185(3), 152–154. <https://doi.org/10.5694/J.1326-5377.2006.TB00504.X>
- Neale, A. V., Dailey, R. K., & Abrams, J. (2010). Analysis of citations to biomedical articles affected by scientific misconduct. *Science and Engineering Ethics*, 16(2), 251–261. <https://doi.org/10.1007/S11948-009-9151-4>
- NSF. (2023). *Publications Output: U.S. Trends and International Comparisons | NSF - National Science Foundation*. <https://nces.nsf.gov/pubs/nsb202333/publication-output-by-region-country-or-economy-and-by-scientific-field>
- Oransky, I. (2022). Retractions are increasing, but not enough. *Nature*, 608(7921), 9. <https://doi.org/10.1038/D41586-022-02071-6>
- Oransky, I., & Marcus, A. (2016). *Two Cheers for the Retraction Boom*. The New Atlantis. <https://www.thenewatlantis.com/publications/two-cheers-for-the-retraction-boom>
- Park, C. ;, Lee, J. Y. ;, & Kwon, J. Y. (2018). How the retracted publications are managed and used? A South Korean case. *STI Conference*, 757–762. <https://hdl.handle.net/1887/65338>
- Rao, T. S. S., & Andrade, C. (2011). The MMR vaccine and autism: Sensation, refutation, retraction, and fraud. *Indian Journal of Psychiatry*, 53(2), 95. <https://doi.org/10.4103/0019-5545.82529>
- Redman, B. K., Yarandi, H. N., & Merz, J. F. (2008). Empirical developments in retraction. *Journal of Medical Ethics*, 34(11), 807–809. <https://doi.org/10.1136/JME.2007.023069>
- Research Areas. (2023). <https://webofscience.help.clarivate.com/en-us/Content/research-areas.html>
- Rosenstock, L., & Lee, L. J. (2002). Attacks on Science: The Risks to Evidence-Based Policy. *American Journal of Public Health*, 92(1), 14. <https://doi.org/10.2105/AJPH.92.1.14>
- Rosokhata, A., Minchenko, M., Khomenko, L., & Chygryn, O. (2021). Renewable energy: a bibliometric analysis. *E3S Web of Conferences*, 250, 03002. <https://doi.org/10.1051/E3SCONF/202125003002>
- Rubbo, P., Helmann, C. L., Bilynievycz dos Santos, C., & Pilatti, L. A. (2019). Retractions in the Engineering Field: A Study on the Web of Science Database. *Ethics & Behavior*, 29(2), 141–155. <https://doi.org/10.1080/10508422.2017.1390667>
- Sacco, D. F., Namuth, A. J., Macchione, A. L., & Brown, M. (2024). Differences in Support for Retractions Based on Information Hazards Among Undergraduates and Federally Funded

- Scientists. *Journal of Academic Ethics*, 1–16. <https://doi.org/10.1007/S10805-024-09505-Y>
- Saikia, P., & Thakuria, B. (2019). Retraction of papers authored by Yuhji Saitoh - Beyond the Fujii phenomenon. *Indian Journal of Anaesthesia*, 63(7), 571–584. https://doi.org/10.4103/IJA.IJA_267_19
- Schmidt, M. (2018). An analysis of the validity of retraction annotation in pubmed and the web of science. *Journal of the Association for Information Science and Technology*, 69(2), 318–328. <https://doi.org/10.1002/ASI.23913>
- Schonhaut, L., Costa-Roldan, I., Oppenheimer, I., Pizarro, V., Han, D., & Díaz, F. (2023). Scientific publication speed and retractions of COVID-19 pandemic original articles. *Revista Panamericana de Salud Pública*, 46, e25. <https://doi.org/10.26633/RPSP.2022.25>
- Serghiou, S., Marton, R. M., & Ioannidis, J. P. A. (2021). Media and social media attention to retracted articles according to Altmetric. *PLoS ONE*, 16(5). <https://doi.org/10.1371/JOURNAL.PONE.0248625>
- Shema, H., Hahn, O., Mazarakis, A., & Peters, I. (2019). Retractions from altmetric and bibliometric perspectives. *Information-Wissenschaft Und Praxis*, 70(2–3), 141–144. <https://doi.org/10.1515/IWP-2019-2006>
- Shepperd, M., & Yousefi, L. (2023). An analysis of retracted papers in Computer Science. *PLOS ONE*, 18(5), e0285383. <https://doi.org/10.1371/JOURNAL.PONE.0285383>
- Sismondo, S. (2008). Pharmaceutical company funding and its consequences: a qualitative systematic review. *Contemporary Clinical Trials*, 29(2), 109–113. <https://doi.org/10.1016/J.CCT.2007.08.001>
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269. <https://doi.org/10.1002/ASI.4630240406>
- Steen, R. G. (2011a). Retractions in the medical literature: how many patients are put at risk by flawed research? *Journal of Medical Ethics*, 37(11), 688–692. <https://doi.org/10.1136/JME.2011.043133>
- Steen, R. G. (2011b). Retractions in the scientific literature: is the incidence of research fraud increasing? *Journal of Medical Ethics*, 37(4), 249–253. <https://doi.org/10.1136/JME.2010.040923>
- Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why Has the Number of Scientific Retractions Increased? *PLOS ONE*, 8(7), e68397. <https://doi.org/10.1371/JOURNAL.PONE.0068397>

- Tang, L., Hu, G., Sui, Y., Yang, Y., & Cao, C. (2020). Retraction: The “Other Face” of Research Collaboration? *Science and Engineering Ethics*, 26(3), 1681–1708. <https://doi.org/10.1007/S11948-020-00209-1>
- Teixeira da Silva, J. A., & Dobránszki, J. (2017). Highly cited retracted papers. *Scientometrics*, 110(3), 1653–1661. <https://doi.org/10.1007/S11192-016-2227-4>
- The Center for Scientific Integrity. (2018). *The Retraction Watch Database*. <http://retractiondatabase.org/>
- Thompson, S. K. (2012). Sampling, Third Edition. *Sampling, Third Edition*. <https://doi.org/10.1002/9781118162934>
- Trikalinos, N. A., Evangelou, E., & Ioannidis, J. P. A. (2008). Falsified papers in high-impact journals were slow to retract and indistinguishable from nonfraudulent papers. *Journal of Clinical Epidemiology*, 61(5), 464–470. <https://doi.org/10.1016/J.JCLINEPI.2007.11.019>
- Umargono, E., Suseno, J. E., & Gunawan, S. K. V. (2020). *K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula*. 121–129. <https://doi.org/10.2991/ASSEHR.K.201010.019>
- van de Velden, M., Iodice D’Enza, A., & Markos, A. (2019). Distance-based clustering of mixed data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3), e1456. <https://doi.org/10.1002/WICS.1456>
- Vuong, Q. H., La, V. P., Ho, M. T., Vuong, T. T., & Ho, M. T. (2020). Characteristics of retracted articles based on retraction data from online sources through February 2019. *Science Editing*, 7(1), 34–44. <https://doi.org/10.6087/KCSE.187>
- Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., Berelowitz, M., Dhillon, A. P., Thomson, M. A., Harvey, P., Valentine, A., Davies, S. E., & Walker-Smith, J. A. (1998). Retracted: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*, 351(9103), 637–641. [https://doi.org/10.1016/S0140-6736\(97\)11096-0](https://doi.org/10.1016/S0140-6736(97)11096-0)
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. <http://www.jstor.org/stable/j.ctt5hhcfr>
- Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645–678. <https://doi.org/10.1109/TNN.2005.845141>
- Zhang, C., Ding, K., & Liu, Z. (2019). *Informetric Analysis on the International Retracted Publication Based on the Web of Science Database*. 472–481. <https://doi.org/10.2991/SSCHD-19.2019.100>

- Zheng, E. Te, & Fu, H. Z. (2022). The Influence of Open Access on the Academic Purification Effect of Retracted Publications: A Causal Inference Analysis. *Proceedings of the Association for Information Science and Technology*, 59(1), 853–855. <https://doi.org/10.1002/PRA2.749>
- Zuckerman, H. (2020). Is “the time ripe” for quantitative research on misconduct in science? *Quantitative Science Studies*, 1(3), 945–958. https://doi.org/10.1162/QSS_A_00065

APPENDIX A: CLASSIFICATION OF VARIABLES

A1. BIOMEDICINE

Table A- 1. Scientific Area Classification

Category	Scientific Area
Arts & Humanities	ARCHITECTURE; ART; ARTS & HUMANITIES OTHER TOPIC; ASIAN STUDIES; CLASSICS; DANCE; FILM, RADIO & TELEVISION HISTORY; HISTORY & PHILOSOPHY OF SCIENCE; LITERATURE; MUSIC; PHILOSOPHY; RELIGION; THEATER
Biomedicine	ALLERGY; ANATOMY & MORPHOLOGY; ANESTHESIOLOGY; AUDIOLOGY & SPEECH-LANGUAGE PATHOLOGY; BEHAVIORAL SCIENCES; CARDIOVASCULAR SYSTEM & CARDIOLOGY; CRITICAL CARE MEDICINE; DENTISTRY, ORAL SURGERY & MEDICINE; DERMATOLOGY; EMERGENCY MEDICINE; ENDOCRINOLOGY & METABOLISM; GASTROENTEROLOGY & HEPATOLOGY; GENERAL & INTERNAL MEDICINE; GENETICS & HEREDITY; GERIATRICS & GERONTOLOGY; HEALTH CARE SCIENCES & SERVICES; HEMATOLOGY; IMMUNOLOGY; INFECTIOUS DISEASES; INTEGRATIVE & COMPLEMENTARY MEDICINE; LEGAL MEDICINE; MEDICAL ETHICS; MEDICAL INFORMATICS; MEDICAL LABORATORY TECHNOLOGY; NEUROSCIENCES & NEUROLOGY; NURSING; NUTRITION & DIETETICS; OBSTETRICS & GYNECOLOGY; ONCOLOGY; OPHTHALMOLOGY; ORTHOPEDICS; OTORHINOLARYNGOLOGY; PATHOLOGY; PEDIATRICS; PHARMACOLOGY & PHARMACY; PHYSIOLOGY; PSYCHIATRY; RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING; REHABILITATION; RESEARCH & EXPERIMENTAL MEDICINE; RESPIRATORY SYSTEM; RHEUMATOLOGY; SUBSTANCE ABUSE; SURGERY; TOXICOLOGY; TROPICAL MEDICINE; UROLOGY & NEPHROLOGY; VIROLOGY
Life Sciences	AGRICULTURE; ANTHROPOLOGY; BIOCHEMISTRY & MOLECULAR BIOLOGY; BIODIVERSITY & CONSERVATION; BIOPHYSICS; BIOTECHNOLOGY & APPLIED MICROBIOLOGY; CELL BIOLOGY; DEVELOPMENTAL BIOLOGY; ENTOMOLOGY; ENVIRONMENTAL SCIENCES & ECOLOGY; EVOLUTIONARY BIOLOGY; FISHERIES; FOOD SCIENCE & TECHNOLOGY; FORESTRY; LIFE SCIENCES BIOMEDICINE OTHER TOPICS; MARINE & FRESHWATER BIOLOGY; MATHEMATICAL & COMPUTATIONAL BIOLOGY; MICROBIOLOGY; MYCOLOGY; PALEONTOLOGY; PARASITOLOGY; PLANT SCIENCES; PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH; REPRODUCTIVE BIOLOGY; SPORT SCIENCES; TRANSPLANTATION; VETERINARY SCIENCES; ZOOLOGY
Physical Sciences	ASTRONOMY & ASTROPHYSICS; CHEMISTRY; CRYSTALLOGRAPHY; ELECTROCHEMISTRY; GEOCHEMISTRY & GEOPHYSICS; GEOLOGY; MATHEMATICS; METEOROLOGY & ATMOSPHERIC SCIENCES; MINERALOGY; MINING & MINERAL PROCESSING; OCEANOGRAPHY; OPTICS; PHYSICAL GEOGRAPHY; PHYSICS; POLYMER SCIENCE; THERMODYNAMICS; WATER RESOURCES

Social Sciences	ARCHAEOLOGY; AREA STUDIES; BIOMEDICAL SOCIAL SCIENCES; BUSINESS & ECONOMICS; COMMUNICATION; CRIMINOLOGY & PENOLOGY; CULTURAL STUDIES; DEMOGRAPHY; DEVELOPMENT STUDIES; EDUCATION & EDUCATIONAL RESEARCH; ETHNIC STUDIES; FAMILY STUDIES; GEOGRAPHY; GOVERNMENT & LAW; INTERNATIONAL RELATIONS; LINGUISTICS; MATHEMATICAL METHODS IN SOCIAL SCIENCES; PSYCHOLOGY; PUBLIC ADMINISTRATION; SOCIAL ISSUES; SOCIAL SCIENCES OTHER TOPICS; SOCIAL WORK; SOCIOLOGY; URBAN STUDIES; WOMEN'S STUDIES
Technology	ACOUSTICS; AUTOMATION & CONTROL SYSTEMS; COMPUTER SCIENCE; CONSTRUCTION & BUILDING TECHNOLOGY; ENERGY & FUELS; ENGINEERING; IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY; INFORMATION SCIENCE & LIBRARY SCIENCE; INSTRUMENTS & INSTRUMENTATION; MATERIALS SCIENCE; MECHANICS; METALLURGY & METALLURGICAL ENGINEERING; MICROSCOPY; NUCLEAR SCIENCE & TECHNOLOGY; OPERATIONS RESEARCH & MANAGEMENT SCIENCE; REMOTE SENSING; ROBOTICS; SCIENCE & TECHNOLOGY OTHER TOPICS; SPECTROSCOPY; TELECOMMUNICATIONS; TRANSPORTATION

A2. MISCONDUCT

Table A- 2. Retraction Reason Classification

Reason	Misconduct	Reason Grouped
Concerns/Issues About Authorship	1	Author issue
Misconduct by Author	1	Misconduct general
Error in Results and/or Conclusions	0	Error general
Investigation by Journal/Publisher	1	Investigation
Investigation by Company/Institution	1	Investigation
Misconduct - Official Investigation/Finding	1	Investigation
Unreliable Results	1	Unreliable
Ethical Violations by Author	1	Misconduct general
Duplication of Data	1	Duplication
Error in Data	0	Error general
Results Not Reproducible	1	Unreliable
Objections by Author(s)	0	Author issue
Retract and Replace	0	Other
Error in Analyses	0	Error general
Error by Third Party	0	Error general
Conflict of Interest	1	Conflict third party/institution
Doing the Right Thing	0	Other
Original Data not Provided	0	Other
Falsification/Fabrication of Results	1	Fabrication/Manipulation
Withdrawn (out of date)	0	Other
Concerns/Issues about Third Party Involvement	1	Conflict third party/institution
Notice - Limited or No Information	0	Other

Temporary Removal	0	Other
Copyright Claims	0	Legal
Manipulation of Images	1	Fabrication/Manipulation
Lack of Approval from Company/Institution	0	Conflict third party/institution
Error in Cell Lines/Tissues	0	Error general
Paper Mill	1	Paper mill
Lack of Approval from Author	0	Author issue
Concerns/Issues About Image	0	Unreliable
Plagiarism of Article	1	Plagiarism
Duplicate Publication through Error by Journal/Publisher	0	Duplication
Complaints about Author	0	Author issue
Euphemisms for Duplication	1	Duplication
Plagiarism of Image	1	Plagiarism
Objections by Third Party	0	Conflict third party/institution
Concerns/Issues about Referencing/Attributions	0	Referencing
Euphemisms for Plagiarism	1	Plagiarism
Contamination of Materials (General)	0	Contamination
Misconduct by Third Party	0	Misconduct general
Plagiarism of Data	1	Plagiarism
Concerns/Issues with Peer Review	0	Peer review
Taken from Dissertation/Thesis	1	Plagiarism
Contamination of Cell Lines/Tissues	0	Contamination
Taken via Peer Review	1	Peer review
Informed/Patient Consent - None/Withdrawn	0	Patient concerns
Notice - Unable to Access via current resources	0	Miscommunication
Legal Reasons/Legal Threats	0	Legal
Criminal Proceedings	1	Legal
Miscommunication by Journal/Publisher	0	Miscommunication
Euphemisms for Misconduct	1	Misconduct general
Concerns/Issues about Human Subject Welfare	0	Patient concerns
Miscommunication by Third Party	0	Miscommunication
Misconduct by Company/Institution	1	Misconduct general
Objections by Company/Institution	0	Conflict third party/institution
Withdrawn to Publish in Different Journal	0	Other
Manipulation of Results	1	Fabrication/Manipulation
Salami Slicing	1	Salami Slicing
Miscommunication by Company/Institution	0	Miscommunication
Complaints about Company/Institution	0	Conflict third party/institution
Complaints about Third Party	0	Conflict third party/institution
Not Presented at Conference	0	Other
Breach of Policy by Third Party	0	Conflict third party/institution
Concerns/Issues About Data	0	Unreliable
Error in Methods	0	Error general

Unreliable Data	1	Unreliable
Upgrade/Update of Prior Notice	0	Other
Lack of IRB/IACUC Approval	0	Patient concerns
Breach of Policy by Author	1	Author issue
Error in Text	0	Error general
Falsification/Fabrication of Data	1	Fabrication/Manipulation
Cites Retracted Work	0	Referencing
Concerns/Issues About Results	0	Unreliable
Bias Issues or Lack of Balance	1	Unreliable
False/Forged Authorship	1	Author issue
Lack of Approval from Third Party	0	Conflict third party/institution
Duplication of Text	1	Duplication
Error in Materials (General)	0	Error general
Investigation by ORI	1	Investigation
Plagiarism of Text	1	Plagiarism
Author Unresponsive	0	Miscommunication
Duplication of Image	1	Duplication
Investigation by Third Party	0	Investigation
Date of Retraction/Other Unknown	0	Other
Duplication of Article	1	Duplication
Error by Journal/Publisher	0	Error general
False Affiliation	1	Fabrication/Manipulation
Updated to Retraction	0	Other
Miscommunication by Author	0	Miscommunication
Falsification/Fabrication of Image	1	Fabrication/Manipulation
Civil Proceedings	1	Legal
Notice - Lack of	0	Other
Publishing Ban	1	Other
Ethical Violations by Third Party	0	Conflict third party/institution
Randomly Generated Content	1	Unreliable
Unreliable Image	1	Unreliable
Rogue Editor	0	Peer review
Notice - No/Limited Information	0	Other
Hoax Paper	1	Misconduct general
Sabotage of Materials	1	Fabrication/Manipulation
Contamination of Reagents	0	Fabrication/Manipulation
Updated to Correction	0	Other
Nonpayment of Fees/Refusal to Pay	0	Legal
No Further Action	0	Other
Error in Image	1	Error general
Fake Peer Review	1	Peer review
Withdrawal	0	Other

APPENDIX B: ADDITIONAL RESULTS

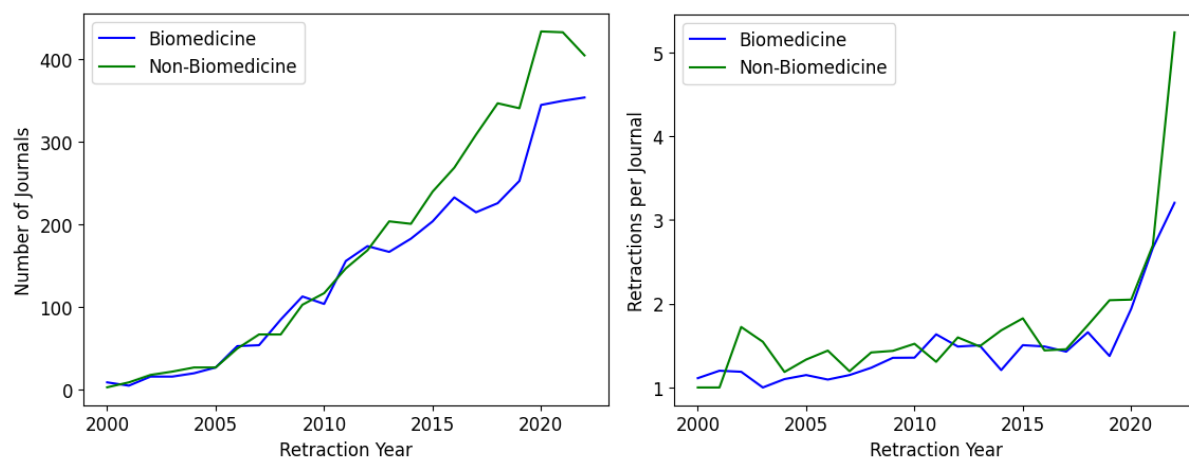


Figure B- 1. Evolution of journals and rate of retraction given retraction year

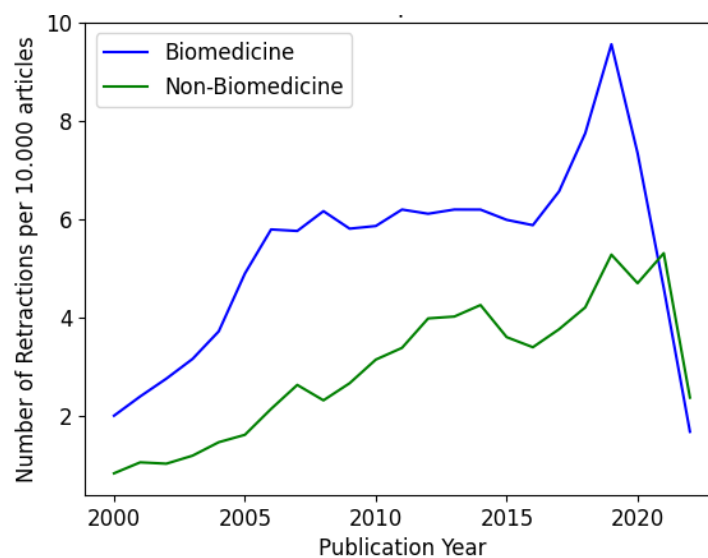


Figure B- 2. Rate of retraction per 10.000 articles

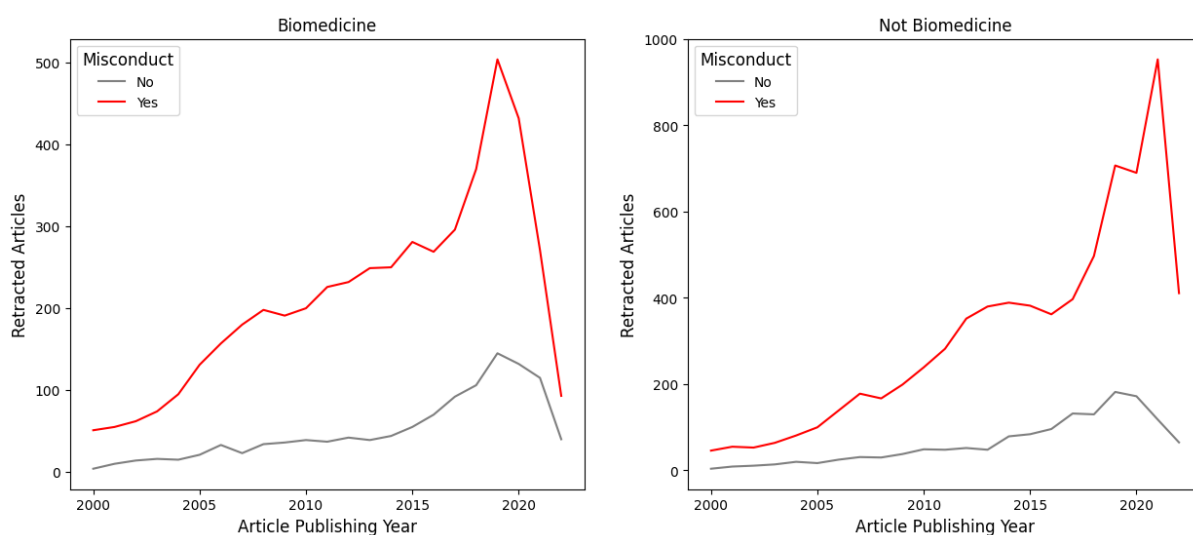


Figure B- 3. Evolution of misconduct given article publishing year

Table B- 1. Reasons for retraction

Reason	Total	Total rate	Biomed	Biomed rate	Non Biomed	Non Biomed rate
Unreliable	4804	33%	2164	36%	2640	31%
Investigation	4767	33%	2148	36%	2619	31%
Duplication	4321	30%	1976	33%	2345	27%
Other	2885	20%	1357	23%	1528	18%
Error general	2578	18%	1115	18%	1463	17%
Peer review	2099	14%	245	4%	1854	22%
Fabrication/ Manipulation	1827	13%	997	17%	830	10%
Plagiarism	1787	12%	607	10%	1180	14%
Author issue	1644	11%	593	10%	1051	12%
Misconduct general	1416	10%	835	14%	581	7%
Conflict third party/institution	1147	8%	451	7%	696	8%
Paper mill	1125	8%	688	11%	437	5%
Referencing	773	5%	149	2%	624	7%
Miscommunication	759	5%	406	7%	353	4%
Patient concerns	373	3%	284	5%	89	1%
Legal	337	2%	138	2%	199	2%
Contamination	87	1%	37	1%	50	1%
Salami Slicing	19	0%	5	0%	14	0%

Table B- 2. Top 15 Authors with most retractions

Author	Number of retractions	Number of Biomedicine Articles	Misconduct
Hironobu Ueshima	113	113	100%
Hiroshi Otake	96	96	100%
Ali Nazari	94	0	100%
Joachim Boldt	78	78	97%
Yoshitaka Fujii	74	74	100%
Yoshihiro Sato	64	57	92%
Jun Iwamoto	64	57	95%
Diederik A Stapel	51	0	100%
Shadi Riahi	47	0	100%
Fazlul H Sarkar	41	32	100%
Wei Zhang	40	29	88%
Lei Wang	39	20	79%
Shahaboddin Shamshirband	38	0	100%
Wei Wang	35	19	77%
Sanjeev Banerjee	34	27	100%

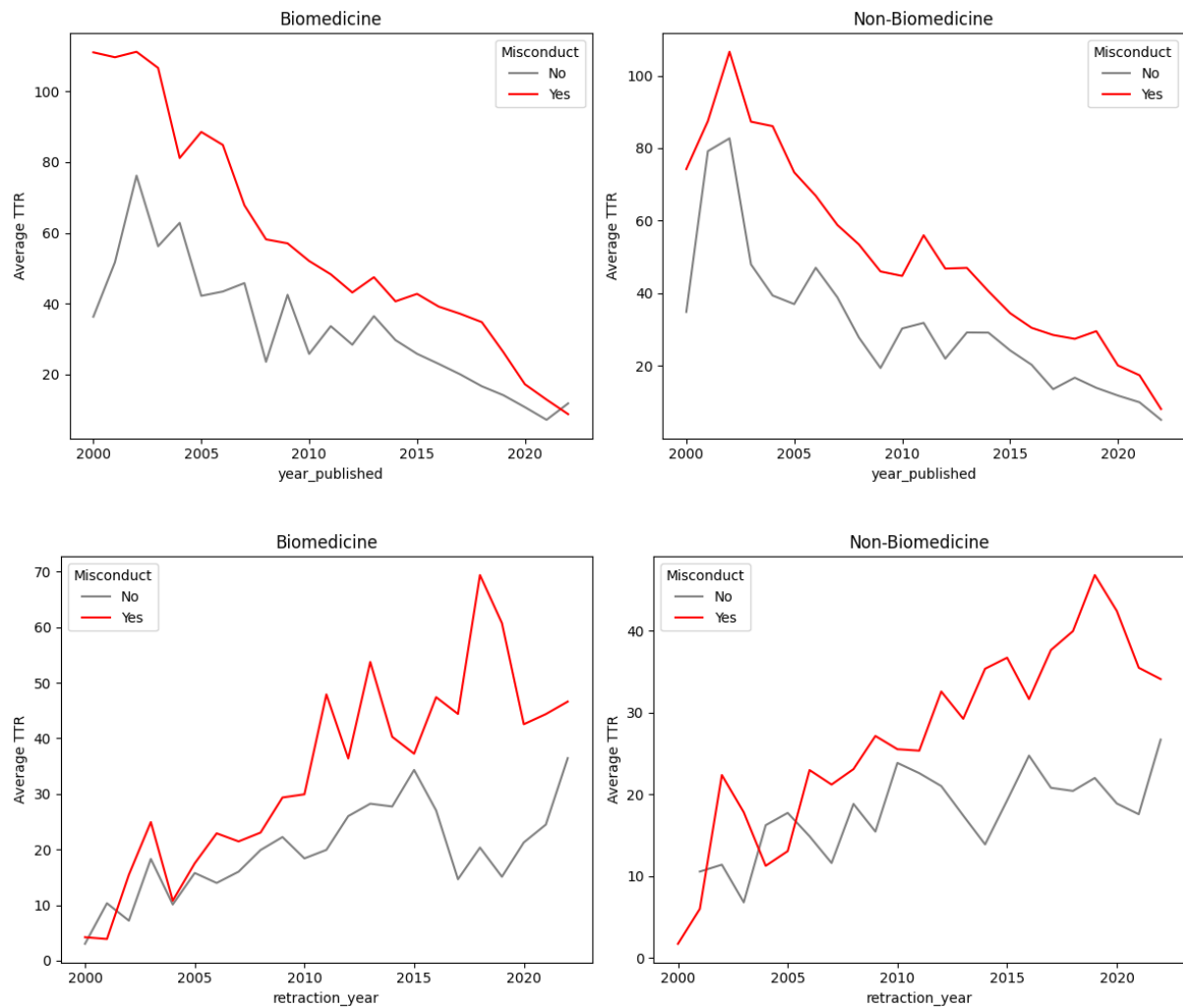


Figure B- 4. Evolution of time to retraction given publishing and retraction year

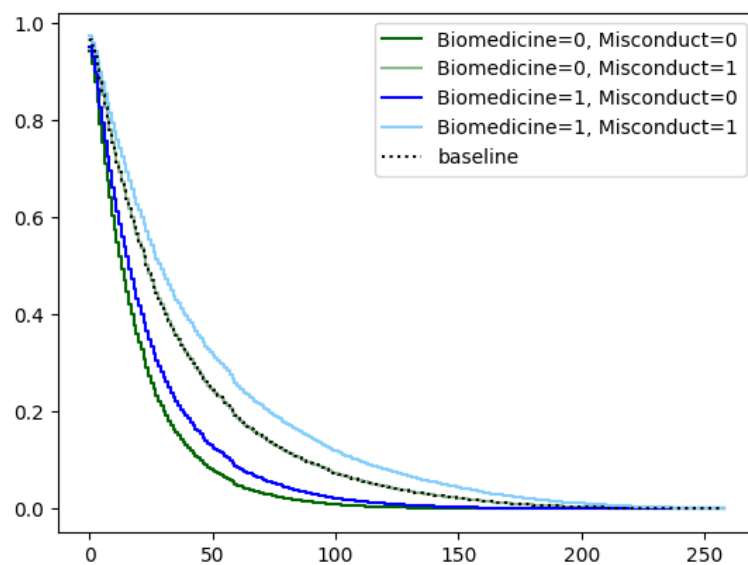


Figure B- 5. Cox Proportional Hazard curve of time to retraction (months)

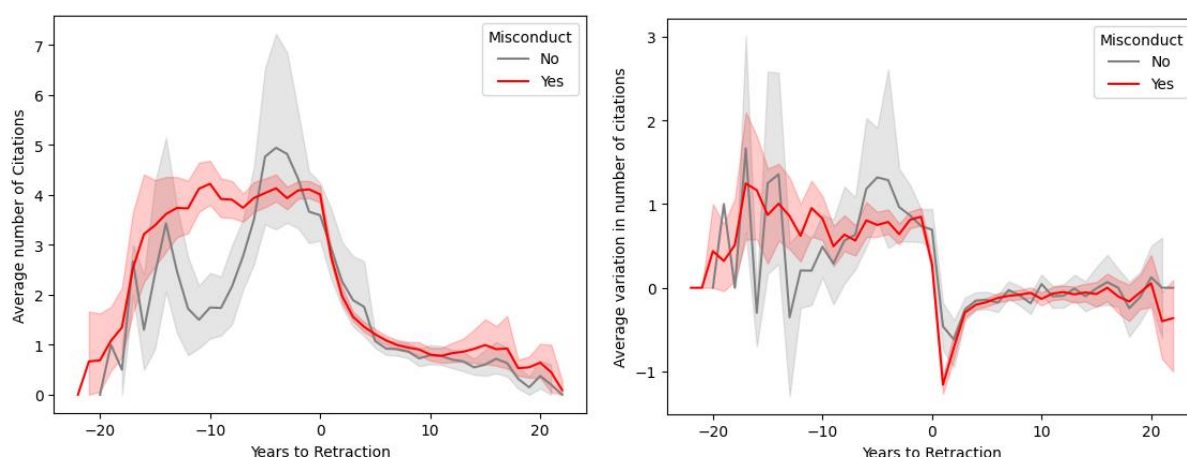


Figure B- 6. Average number and variation of citations according to time to retraction

Table B- 3. References of most cited retracted articles

Article	Reference
Tracheobronchial transplantation with a stem-cell-seeded bioartificial nanocomposite: a proof-of-concept study	Jungebluth, P., Alici, E., Baiguera, S., Blanc, K. le, Blomberg, P., Bozóky, B., Crowley, C., Einarsson, O., Grinnemo, K. H., Gudbjartsson, T., Guyader, S. le, Henriksson, G., Hermanson, O., Juto, J. E., Leidner, B., Lilja, T., Liska, J., Luedde, T., Lundin, V., ... MacChiarini, P. (2011). Tracheobronchial transplantation with a stem-cell-seeded bioartificial nanocomposite: a proof-of-concept study. <i>The Lancet</i> , 378(9808), 1997–2004. https://doi.org/10.1016/S0140-6736(11)61715-7
Selective killing of cancer cells by a small molecule targeting the stress response to ROS	Raj, L., Ide, T., Gurkar, A. U., Foley, M., Schenone, M., Li, X., Tolliday, N. J., Golub, T. R., Carr, S. A., Shamji, A. F., Stern, A. M., Mandinova, A., Schreiber, S. L., & Lee, S. W. (2011). Selective killing of cancer cells by a small molecule targeting the stress response to ROS. <i>Nature</i> 2011 475:7355, 475(7355), 231–234. https://doi.org/10.1038/nature10167
Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study	Cai, Q., Yang, M., Liu, D., Chen, J., Shu, D., Xia, J., Liao, X., Gu, Y., Cai, Q., Yang, Y., Shen, C., Li, X., Peng, L., Huang, D., Zhang, J., Zhang, S., Wang, F., Liu, J., Chen, L., ... Liu, L. (2020). Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study. <i>Engineering</i> , 6(10), 1192–1198. https://doi.org/10.1016/J.ENG.2020.03.007
An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus	Voinnet, O., Rivas, S., Mestre, P., & Baulcombe, D. (2003). Retracted: An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. <i>The Plant Journal</i> , 33(5), 949–956. https://doi.org/10.1046/J.1365-313X.2003.01676.X

Primary Prevention of Cardiovascular Disease with a Mediterranean Diet	Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M.-I., Corella, D., Arós, F., Gómez-Gracia, E., Ruiz-Gutiérrez, V., Fiol, M., Lapetra, J., Lamuela-Raventos, R. M., Serra-Majem, L., Pintó, X., Basora, J., Muñoz, M. A., Sorlí, J. v., Martínez, J. A., & Martínez-González, M. A. (2013). Primary prevention of cardiovascular disease with a mediterranean diet. <i>Zeitschrift Fur Gefassmedizin</i> , 10(2), 28. https://doi.org/10.1056/NEJMOA1200303
Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19	Mehra, M. R., Desai, S. S., Kuy, S., Henry, T. D., & Patel, A. N. (2020). Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. <i>New England Journal of Medicine</i> , 382(25), 102. https://doi.org/10.1056/NEJMOA2007621
Visfatin: A Protein Secreted by Visceral Fat That Mimics the Effects of Insulin	Fukuhara, A., Matsuda, M., Nishizawa, M., Segawa, K., Tanaka, M., Kishimoto, K., Matsuki, Y., Murakami, M., Ichisaka, T., Murakami, H., Watanabe, E., Takagi, T., Akiyoshi, M., Ohtsubo, T., Kihara, S., Yamashita, S., Makishima, M., Funahashi, T., Yamanaka, S., ... Shimomura, I. (2005). Visfatin: A protein secreted by visceral fat that Mimics the effects of insulin. <i>Science</i> , 307(5708), 426–430. https://doi.org/10.1126/SCIENCE.1097243
Apoptosis and Molecular Targeting Therapy in Cancer	Hassan, M., Watari, H., Abualmaaty, A., Ohba, Y., & Sakuragi, N. (2014). Apoptosis and Molecular Targeting Therapy in Cancer. <i>BioMed Research International</i> , 2014(1), 150845. https://doi.org/10.1155/2014/150845
Spontaneous Human Adult Stem Cell Transformation	Rubio, D., Garcia-Castro, J., Martín, M. C., de La Fuente, R., Cigudosa, J. C., Lloyd, A. C., & Bernad, A. (2005). Spontaneous Human Adult Stem Cell Transformation. <i>Cancer Research</i> , 65(8), 3035–3039. https://doi.org/10.1158/0008-5472.CAN-04-4194
Purification and ex vivo expansion of postnatal human marrow mesodermal progenitor cells	Reyes, M., Lund, T., Lenvik, T., Aguiar, D., Koodie, L., & Verfaillie, C. M. (2001). Purification and ex vivo expansion of postnatal human marrow mesodermal progenitor cells. <i>Blood</i> , 98(9), 2615–2625. https://doi.org/10.1182/BLOOD.V98.9.2615

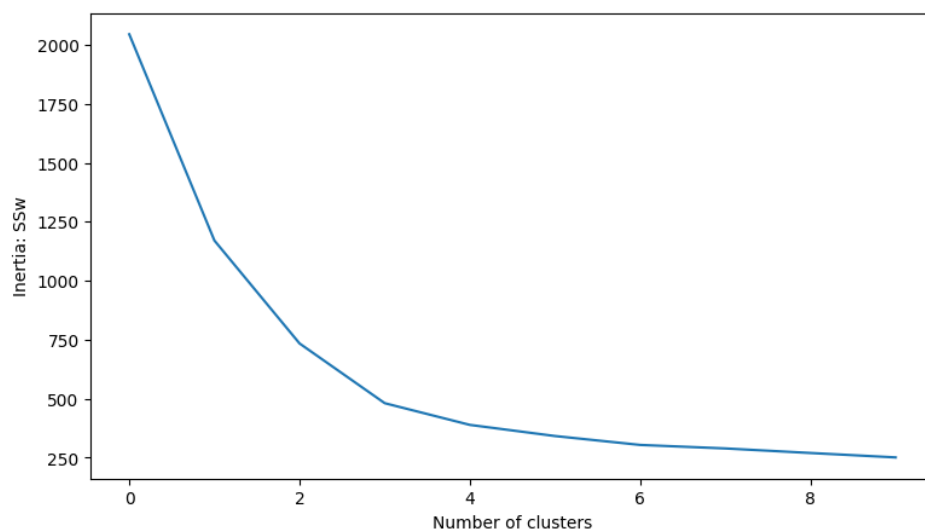


Figure B- 7. Inertia plot for choice of number of clusters

ANNEX A: SURVIVAL MODELS

Survival models are typically used in medical research to measure the effect of a treatment by assessing the number of surviving patients over a period of time (Goel et al., 2010). More broadly, they can be used to estimate the survival time (lifespan from starting point to event of interest) of a particular population and in a defined study period. They measure not only the time difference of the individuals until a particular event, but also if that event happened in the period of study.

Two types of models were implemented. The Kaplan-Meier is a univariate approach, while the Cox Proportional Hazards model is multivariate. Both are always delineated by a starting event and the end event which we aim to estimate its occurrence. Then, the period following the starting event is divided into several time intervals, and the number of end occurrences (e.g. deaths) that happen in each time frame are noted. A survival function, S_t , is created to map the probability of survival at any given time t . The hazard function, $h(t)$, indicates the rate at which an event-free observation will instantaneously experience the event. This means that an individual has S_t probability of having survived until time t and, if they survive until that point, they have an $h(t)$ probability of suffering the event immediately after.

The Kaplan-Meier estimator works under three assumptions:

1. At any time, all censored individuals have the same probability of survival as the observed individuals. Censored individuals are those that stop being observed before the end of the study period, or do not register the event until the end of the study period. These subjects only register partial information, as we will not be able to ascertain the date of the occurrence of the end event.
2. Individuals whose observation starts later have the same probability as the subjects that started being observed earlier (in relation to the starting event).
3. The event happens at the time specified in the observation.

The conditions of the bibliometric databases are not alike to controlled trials, since the revision of all the articles is not done at regular time intervals, but as information becomes available. This can either be as the document is published/modified in the journal, or at a later date. So, censored observations in this case are the articles whose status have not been thoroughly updated at the date of the study. The second assumption is mitigated through the fact that the databases about articles and retractions are in continual monitoring and update. For this reason, it is expected that all articles are observed from the moment of their starting event (the date of the publishing of the article), up until the date of the retraction (end event). Regarding the third assumption, even if the retraction has a lag until it is registered in databases, the date of the retraction is registered as the retraction publishing date, available in the metadata of the document. This ensures that there is no bias stemming from misalignment of the date of the real occurrence and of the observation of the event.

Since the Kaplan-Meier estimate is a non-parametric statistic (Clark et al., 2003), no assumption is made on the distribution of the data. It is also known as the product limit estimate, since the survival curve is created by computing probabilities of occurrence of the event at several points in time, and multiplying those successive probabilities by the earlier ones to get the final estimate. The survival function is given by formula 4.

$$(4) S_t = P(T > t) = S_{t-1} \left(1 - \frac{d_j}{n_j} \right)$$

Where T is a random variable denoting the time of the event, t is any specified time point, and $P(T > t)$ represents the probability of not experiencing the event until, and including, time t . In $t = 0$, the survival probability is 1 ($S_{t=0} = 1$). In subsequent time periods, the probability of survival at time t is calculated from the probability of survival the previous period, S_{t-1} , times the number of events at t (d_j), divided by the number of patients alive just before t (n_j).

The hazard function is obtained through the S_t , as formula 5 indicates.

$$(5) h_t = -\frac{d}{dt} [\log(S_t)]$$

Since it is not simple to estimate the quantity h_t , it is common to use the cumulative hazard H_t . It is the integral of the hazard, corresponding to the symmetric of the log-survivor curve, $H_t = -\log(S_t)$. It represents the number of events that would be expected for each individual by time t if the event were a repeatable process.

In the case of time to retraction, the survival probability is calculated for each time interval as the number of articles who have not yet been retracted divided by the number of articles not retracted at the beginning of the time interval. The estimates are typically observed in graphical form, showing the evolution of the survival probability at each time interval t . It is drawn as a step function called survival curve where, between events, the probability maintains constant.

The Kaplan-Meier estimator has two important limitations. Since it is a univariate analysis, so it does not allow for the consideration of multiple factors simultaneously. This means that it is not possible to obtain the relative risk or related confidence interval and compare the survival of different subgroups of individuals. It is, however, possible to compare the survival curve of different groups (categorical variables) by plotting them into the same graph. Additionally, this estimator requires data discretization of variables, which means that it is not possible to calculate the difference in relative risk for incremental changes of a certain variable (Abd Elhafeez et al., 2021).

The Cox Proportional Hazards model is able to discern the effect of several predictors on the survival probability on an individual. It uses the hazard function to understand the relationship between the predictors and the time-to-event (George et al., 2014). It assumes that the

predictors have a multiplicative effect on the hazard, and that this effect does not change over time.

The hazard regression model is given by formula 6:

$$(6) \ h(t) = h_0(t) \times \exp[b_1x_1 + b_2x_2 + \dots + b_kx_k]$$

Where $h_0(t)$ is the baseline hazard at time zero, equivalent to all predictors being set to zero. The parameters are interpreted as the estimated increase in the hazard of the event happening to an individual with risk factor x , compared to other individuals without it (Abd Elhafeez et al., 2021).

ANNEX B: NEGATIVE BINOMIAL REGRESSION

When performing regressions, it is important to consider the type of the target variable. Using an OLS model on a count variable, would lead to negative or non-integer predicted values, which do not reflect the reality. It may also lead to inefficient estimates, by not accounting for the heteroskedastic nature of event counts (Ghazal & Zulkhibri, 2015). Methods like the Poisson regression deal with these factors directly by assuming that y given x follows a Poisson distribution. Using a *quasi*-maximum likelihood estimator, this assumption of distribution is unnecessary (Wooldridge, 2010). Unfortunately, these models assume that the variance-mean ratio is a one. In cases where the variance is higher than the mean, there is overdispersion.

The Negative Binomial regression is a generalization of the Poisson model that is able to deal with overdispersion (Greene, 2003). It introduces an individual unobserved effect to the conditional mean, as follows

$$(7) \mu_{it} = \exp(X_{it}\beta) \times c_i = \exp(X_{it}\beta + \eta_i)$$

Where $X_{it}\beta$ is the set of exogenous variables that control the count variable μ for individual i at time t . The expression has the consequence of capturing the effect of the variables on the log count and not the count itself. To obtain the variation of the count for a given variable, one needs to apply the exponential to the coefficient.

ANNEX C: PROPORTIONAL STRATIFIED RANDOM SAMPLING

In this method, the population is partitioned into *strata*, and a probabilistic sample is selected independently for each of those partitions (Thompson, 2012). The aggregation of the samples selected from each *stratum* (h) composes the global sample, from which the parameters are estimated. The groups should be exhaustive and mutually exclusive, so every element in the population must appear once and only once in the *strata*. The specific case of proportional allocation refers to the dimensionality of the selected sample being proportional to the dimensionality of the *stratum*. Formula 8 for obtaining the sample size, n , is the following:

$$(8) \quad n = \frac{z_{1-\alpha/2}^2 N S_{intra}^2}{d^2 N + z_{1-\alpha/2}^2 S_{intra}^2}$$

Where $z_{1-\alpha/2}$ corresponds to the quantile from the Normal distribution, given the confidence level chosen ($\alpha = 5\%$). N is the population size, corresponding to the number of corrections associated with pre-retraction citations. d is the desired precision, which will be 0.15 in this context. Finally, $S_{intra}^2 = \sum_h \frac{N_h}{N} S_h^2$, which is the weighted average of the variance in the strata. Since the variance of the parameter is not known, and no historical information is available, it was opted to use the pessimistic approach. The worst-case scenario for the variance of a proportion is when it is its highest value: 0.5. This is applied to all strata.

To allocate the sample to each stratum, we use formula 9:

$$(9) \quad n_h = \frac{n}{N} N_h$$

A random sample is then selected from each strata with its corresponding dimension. To select the samples, only corrections with available DOI information were considered. Then, each correction was manually verified for any mention of retracted articles as reason for correction. To obtain the final estimate of the proportion, the following formula was applied:

$$(10) \quad p = \sum_{h=1}^H \frac{N_h}{N} p_h$$

To understand the quality of estimation, the precision is calculated. The first step is to calculate the variance of the estimator.

$$(11) \quad \hat{V}(\hat{p}) = \frac{\bar{s}^2}{n} - \frac{S_{intra}^2}{N}, \text{ onde } \bar{s} = \sum_h \frac{N_h}{N} s_h$$

From this value, the estimated absolute precision is given by:

$$(12) \quad d = z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{p})}$$

ANNEX D: CLUSTER ANALYSIS

Cluster Analysis are used to create natural groupings of the sample and aim to detect the underlying structure of the population (Xu & Wunsch, 2005). The specific implementation chosen was a combination of the Gower distance with the K-Medoids algorithm.

The Gower distance is characterised by the application of different similarity measures according to the datatype (Gower, 1971). For numerical and ordinal, it computes one minus the normalised Manhattan distance, with the particularity of ordering data beforehand for the latter case. For categorical variables, matching observations had a similarity of 1 and differing ones had value 0. The similarity between two individuals i and j are formalised below.

$$(13) S_{ij} = \frac{\sum_{k=1}^v S_{ijk}}{\sum_{k=1}^v \delta_{ijk}}$$

If k is quantitative:

$$(14) S_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k}$$

If k is qualitative or dichotomous:

$$(15) S_{ijk} = \begin{cases} 1 & \text{if } x_{ik} = x_{jk} \\ 0 & \text{if } x_{ik} \neq x_{jk} \end{cases}$$

Where R_k represents the range of values the variable k takes, and $\delta_{x,y}$ represents the number of valid variables, i.e., with non-null values in both compared observations. It is also possible to add weights to the variables, but since that was not applied in this context, the formulas show the simplified version of the distance.

Clustering algorithms typically work based on the distance between observations. From the Gower similarity measure it is intuitive that $1 - S_{ij}$ will provide the distance between observations. Once this measure is defined, it is possible to apply a clustering algorithm. van de Velden et al. (2019) suggests the implementation of the K-Medoids algorithm for its robustness in comparison to the K-Means, and its ability to handle the chosen distance metric.

The K-Medoids algorithm partitions the dataset into a pre-determined number of clusters according to the natural assignment of data point to a cluster centre, known as medoid (Jin & Han, 2011). It starts by randomly assigning k observations to be the initial medoids. Then, each observation is assigned to the medoid which minimises the distance between the observation and the medoid, forming k distinct groups. The medoid of each group is reassigned to be the most centrally located observation, i.e., the one that minimises the sum of distances to other points within the cluster. Followingly, an iterative process starts where the distances between observations and medoids are computed and the medoids readjusted. The process ends when

one of the following termination conditions is met: either the number of maximum iterations is reached, or the medoids stabilised.

Since the K-Medoids algorithm needs a pre-defined number of clusters, the elbow method was used to determine the best number of clusters. It is a popular technique due to its intuitive interpretation and visual representation (Umargono et al., 2020). Essentially, it evaluates the performance of the algorithm for different number of clusters and chooses the point at which the gains in performance become marginal (Cui, 2020). It is expected that the performance increases as the number of clusters increases, because the within cluster variance will decrease with less observations per cluster. However, at a certain point, the addition of new clusters will separate observations that are already very similar in nature, in which case it defeats the purpose of clustering. The sum of distances to the closest medoid (*inertia_* attribute in the python package *sklearn_extra.cluster.KMedoids*) are compared for each number of clusters, and visualised in a line plot. At the point with the most extreme angle difference between the previous and subsequent number of clusters, is the optimal choice.

