Journal retractions in oncology – a bibliometric study

Pan Pantziarka Corresp., [1, 2] , **Lydie Meheus** [1]

[1] Anticancer Fund, Brussels, Belgium
[2] The George Pantziarka TP53 Trust, London, United Kingdom

Corresponding Author: Pan Pantziarka
Email address: anticancer.org.uk@gmail.com

Background. A number of recent high-profile cases have focused attention on scientific misconduct and other problematic issues with the peer review process. The retraction of journal publications is an important part of the scientific publishing process that serves to remove flawed articles, (including, but not limited to, fraudulent results), from the literature. To date there have been few formal studies of journal retractions in the area of oncology. **Methods.** This article outlines the results of a bibliometric study of journal retractions from 1983 to 2018. **Results.** Analysis shows that article lifetime – that is the time period from initial publication to ultimate retraction – has decreased in recent years. It also shows that retraction rate has also increased over the same period. The causes and context of these trends are discussed and reference made to the dangers of scientific misconduct in oncology.

Journal Retractions in Oncology – A Bibliometric Study

Pan Pantziarka [1, 2], Lydie Meheus [2]

[1] The George Pantziarka TP53 Trust, London KT1 2JP, UK
[2] Anticancer Fund, Brussels, 1853 Strombeek-Bever, Belgium

Corresponding Author:
Pan Pantziarka
Email address: anticancer.org.uk@gmail.com

Abstract

Background. A number of recent high-profile cases have focused attention on scientific misconduct and other problematic issues with the peer review process. The retraction of journal publications is an important part of the scientific publishing process that serves to remove flawed articles, (including, but not limited to, fraudulent results), from the literature. To date there have been few formal studies of journal retractions in the area of oncology.

Methods. This article outlines the results of a bibliometric study of journal retractions from 1983 to 2018.

Results. Analysis shows that article lifetime – that is the time period from initial publication to ultimate retraction – has decreased in recent years. It also shows that retraction rate has also increased over the same period. The causes and context of these trends are discussed and reference made to the dangers of scientific misconduct in oncology.

Introduction

In recent years there has been an increased focus on the retraction of scientific articles due to scientific misconduct – including cases where authors have been shown to have faked data or doctored images, committed plagiarism or other acts of misconduct. Websites such as RetractionWatch and PubPeer have served to highlight cases of misconduct or to raise expressions of concern leading to further investigations into suspected cases of fraud (Didier & Guaspare-Cartron, 2018).

There has been a concern that the number of retracted articles has been increasing in recent years (Steen, 2011; Steen, Casadevall & Fang, 2013). This is particularly important in the biomedical field where flawed publications may serve to mislead other scientists, clinicians and patients. It is notable that two prolific authors of retracted publications, Fazlul H. Sarkar and Bharat B. Aggarwal, were active in the field of oncology and both focused on alternative cancer

treatments (Dyer, 2016; Offord, 2018). Retracted publications are potentially damaging even in cases where there is no ill intent and the retractions are due to errors rather than to misconduct.

This paper focuses on article retractions specifically in oncology. In addition to characterising the landscape of article retractions in this area in recent years, an attempt is made to put this into the context of the growth of the oncology literature over the same period. Secondly, this article addresses the question of whether the period of time between publication and retraction shows any trend over time. Finally, the question of the influence, in terms of citations, of these articles is also addressed.

Methods

A bibliometric analysis of retracted oncology publications from the Pubmed database was performed.

Search Criteria

Pubmed uses the MESH term “Retraction of Publication” (MeSH Unique ID: D016440) to identify publications announcing the retraction of a publication (i.e. retraction notices). Oncology publications, from all areas of cancer research, including pre-clinical and clinical studies, were included by using the search terms "neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]. No restrictions were placed on time period and no other filters were applied, therefore yielding the final search string of:
"Retraction of Publication" [Publication Type] and ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]).

Data Processing

Each Pubmed retraction notice, denoted by a unique PMID, may include one or more PMIDs identifying the original article being retracted. The PMIDs for these retracted articles were therefore extracted and used to create the corpus of retracted articles for further analysis. Relevant fields for each retracted article included PMID, DOI, PMID of associated retraction notice, article title, abstract, journal, authors and author affiliations. Each record included the date of publication of the article in a journal, the specific date of publication for the article and the date on which the article record was completed in Pubmed. The earliest of these three dates was used as the date on which the article became available in the literature. A similar process established the date on which the retraction notice was first made available in the literature. The difference, in days, between the publication date of the retraction notice and the original article was calculated as the ‘lifetime’ of the original article.

Citations

Citation data is available from a number of different data repositories, each with a different methodology and yielding different numbers of citations per article. Citation counts for each original article were downloaded from three sources: Google Scholar, Mendeley and Clarivate Web of Science. Where an article had no citations listed in one of these sources the value was recorded as zero. A final citation score was calculated as the mean of the non-zero citation counts from the three citation sources.

Authors

The list of authors for each retracted article was also extracted and processed to correct for alternative forms (for example with and without middle initial, alternative hyphenation and so on). The cleansed list was then used to generate a list of authors and the count of the number of retracted articles they had authored. Fractional authorship was not calculated, each author was credited a count of one if his or her name appeared as an author of a retracted article.

Global Oncology Corpus

In order to place the pattern of oncology retractions in context the Pubmed database was searched using the terms: ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]). The number of articles per year was downloaded and stored.

Results

Retraction Notices

A Pubmed search of retraction notices was carried out on October 25th 2018, yielding a total of 627 unique retraction notices. Of these 53 were ‘in situ’ retractions in that the PMID of the original article was also used for the retraction notice. The majority, 574, had unique PMIDs and listed the PMIDs of the retracted articles separately. Six of the latter were notices of multiple retractions in which two or more original articles were announced as being retracted. In total the PMIDs of 591 original articles were extracted from the data in the retraction notices.

Retracted Articles

A dataset of 591 retracted articles was constructed as described in the Methodology. The dataset included retracted articles originally published in the period 1983 to 2018, although in eight of those years no articles that were published have since been retracted (1984-88, 1990, 1993 and 1999). The annual distribution of retracted articles is shown in Figure 1.

For the 28 years with retractions the mean number of retracted articles per year was 21.1, the median 14.6, standard deviation 21.4 and interquartile range(IQR) 36.3 – 2.0.
The article lifetime, defined as the number of days between publication of the original article and the number of days between publication of the retraction notice is shown in Figure 2.

For the 28 years with retractions the mean lifetime per retracted article per year was 1294 days, the median 780, the standard deviation 1280 and the IQR 1938 – 128. Given that the period 1983 – 1999 has so few retracted articles, and that there is a median article lifetime of more than 2 years, we can perhaps focus on the period 2000 – 2016 to assess whether the extent of the apparent trends in article count and article lifetime.

Figure 3A shows the mean of the retracted article counts, and Figure 3B shows the associated lifetimes in retracted oncology articles listed in Pubmed for the period 2000 – 2016. Linear trend lines are also shown.

To place these figures in context Figure 4A shows the increase in oncology articles in Pubmed between 2000 and 2016, exhibiting a very strong linear trend. Figure 4B shows the number of retractions per 10k articles over the same period. We can conclude that it is not just the absolute number of retractions that has increased; it is also the rate of retractions in the oncology literature as a whole. There is also clearly a negative correlation between the article lifetimes in Figure 3B and the retraction rate in Figure 4B, a relationship that is statistically significant ($P < 0.008$).

Authors

In contrast to the downward trend in article lifetime, there is no discernible change in the mean number of authors per retracted paper. Figure 5 shows the distribution of author counts in the entire period, including (in red), the mean for the period 2000 – 2016.

Across the entire period there are 2994 authors, with the vast majority (87%) only authoring one retracted article, and 9% authoring two retracted articles, Figure 6A. The 10 most prolific authors have 196 articles between them, with the most three prolific having 40, 33 and 29 retractions each respectively. Figure 6B shows the distribution of the number of authors per article, showing a peak at the median of 6, and also a small number of articles (3%) with large (>15) author counts.

Citations

The article citation count is the basis of a number of key bibliometric measures, including the h-index (Hirsch, 2005). The citation counts of the retractions in this data set are highly skewed, as shown in Figure 7. The mean citation count is 43.5, with a median of 16, IQR of 45.3 – 5.7 and standard deviation of 86.8. Intuitively one would expect to see a relationship between article lifetime and citation count, and indeed there is a significant relationship between the shorter lifetime and the lower citation rates over the entire period and for the period 2000 – 2016, (both $P < 0.008$).

Journals

The retracted articles in this corpus appeared in 231 journals over the entire period. As shown in Figure 8 the relationship is skewed. The majority of journals published one or two articles which were retracted over the entire period. However, 8% of journals had published more than 5 retracted articles. The mean number of retracted articles published per journal over the entire period was 2.6, the median 1.0, IQR 1.0 – 1.0 and the standard deviation 4.3. One notable characteristic of the journals with the highest numbers of retractions is that they had published some of the authors with the highest number of retractions – and that the serial retractions of these authors contribute strongly to the skewed relationship.

Discussion

This work has focused on the pattern of retractions in the field of oncology and has not addressed the issue of the causes of these retractions. Fang et al assessed the causes of article retractions in biomedicine and the life-sciences and determined that 67.4% of retractions were due to misconduct in one form or another, and only 21.3% due to error (Fang, Steen & Casadevall, 2012). Bozzo and colleagues focused purely on retractions in oncology and determined that research misconduct was the cause of 61% of retractions (Bozzo et al., 2017). Certainly in this work many of the highest scoring authors of retracted papers have been found guilty of research misconduct that has occurred over many years. The serial retractions of the work of these authors can be judged by the skewed distribution shown in Figure 8. It is in the context of this pattern of long-term serial offending that the question of the temporal trends in oncology retractions should be addressed.

The number of retracted articles in oncology has shown a sustained annual increase during the period 1983 – 2018, as shown in Figure 1. For the period 2000 – 2016, as shown in Figure 3A, there is a strong linear trend representing an average annual growth rate of 45.7%. Over the same period the oncology literature also showed a sustained annual increase in the number of publications, as shown in Figure 4A, representing an average annual growth rate of 5.8%. The data shows, in Figure 4B, that the number of retractions as a proportion of the oncology literature has also increased. It is a fair assumption therefore that the number of retractions is likely to continue its upward trajectory in the near term.

Over the same period there has also been a very clear downward trend in article lifetime, as shown in Figure 2 for the entire period, and in Figure 3B for the period 2000 – 2016. This is a very positive finding, particularly as it suggests that erroneous or fraudulent articles are being removed from the literature more quickly despite the increases in the number of cancer-related publications. One consequence should be that these retracted articles are removed before they accrue significant numbers of citations. This assumption is supported by the data shown in Figure 7, which shows a decreasing number of citations per retracted article. One may ask whether these articles are highly cited even if the average citation rate is falling. Clearly there are some articles which have high citation rates, for example Figure 7 shows clear peaks in 1989 and 2002. A recent study analysed citation rates in a number of scientific fields in order to assess the relative patterns of citations across different areas of study (Patience et al., 2017). For oncology the most highly cited articles in the period 2010 – 2014 had a mean of 260 citations. In our dataset only 18/591 (3%), exceeded this value – the most recent of which dated from 2011. For the last two years for which we have reasonably complete data, (given the median article lifetime of 780 days), 2015 and 2016, the highest number of citations was 85 and the mean was 8.9.

While there are clear trends in article counts, retraction rates and article lifetimes there are few clear trends in other areas. The number of authors per retracted article has remained fairly consistent with a median of six, as shown in Figure 5 and Figure 6. Given the relative ease with which a sole author could commit fraud or other misconduct, it is notable that less than 4% of

articles of have single authorship. In terms of the data on journals, as shown in Figure 8, the vast majority of journals (76%) have published only one or two retracted articles. A number of journals score high in this area primarily because they have published articles by some of the most prolific authors of retracted articles.

Why are we seeing more retractions in the oncology literature? One possible explanation lies in the 'many eyes hypothesis' which posits group size effects that lead to increases in anti-predator vigilance among animals –at its most simple it states that as group size increases there are more eyes scanning the environment for predators (Lima, 1995) . A variant of this hypothesis, known as Linus's Law, has become popular in the field of open-source software development and it is often stated as "given enough eyeballs, all bugs are shallow" (Wikipedia contributors, 2018). Certainly the growth in the published literature, Figure 4A, may be viewed as indirect evidence of increased group size. We may also view initiatives such as PubPeer and RetractionWatch, including the newly released RetractionWatch database, as mechanisms conducive to increased vigilance on the part of the scientific community. We make no comment as to the implication that scientific fraudsters are predators seeking advantage by attacking the integrity of the scientific process.

A very good illustration of the dangers of fraudulent scientific publications is the case of Nobuto Yamamoto and GcMAF, of whom three retracted publications are included in this dataset (Ugarte, Bouche & Meheus, 2014). These retracted papers were used as foundational texts by sophisticated fraudsters selling fake cures for cancer, autism and HIV (BBC, 2018). One may note that while the courts have been active in closing down this sophisticated and dangerous operation, a number of papers by the fraudsters have yet to be retracted and remain in the literature despite legitimate concerns about the veracity of the claims made by the authors.

This work has clear limitations. It does not attempt to address the causes of the retractions in this dataset. It is based on a single source of data, Pubmed, and therefore almost certainly misses relevant publications which are not indexed there. Within this dataset there are retraction notices which do not include details of the original publications. The practice of reusing a PMID of the original article for the retraction notice, which we have termed 'in situ retractions', makes automated bibliometric analysis difficult.

Conclusions

The data presented in this paper indicates that the number of retracted articles in oncology is on an upward trajectory, as is the share of retracted articles of the overall oncology literature. However, this increase in retractions is accompanied by a trend of lowered citation rates and lower article lifetimes. This positive trend suggests that the oncology community is becoming more active in identifying suspect articles. Given the potential for harm that arises from erroneous and/or fraudulent articles these positive trends are to be welcomed.

Acknowledgements

References

BBC. 2018.GcMAF: Millions made from sale of “cancer cure.” Available at https://www.bbc.co.uk/news/world-europe-guernsey-46294021 (accessed November 26, 2018).

Bozzo A., Bali K., Evaniew N., Ghert M. 2017. Retractions in cancer research: a systematic survey. Research integrity and peer review 2:5.

Didier E., Guaspare-Cartron C. 2018. The new watchdogs’ vision of science: A roundtable with Ivan Oransky (Retraction Watch) and Brandon Stell (PubPeer). Social studies of science 48:165–167.

Dyer O. 2016. Researcher suing PubPeer was found culpable of misconduct, court documents show. BMJ (Clinical research ed.) 355:i6337.

Fang FC., Steen RG., Casadevall A. 2012. Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109:17028–17033.

Hirsch JE. 2005. An index to quantify an individual’s scientific research output. Proceedings of the National Academy of Sciences of the United States of America 102:16569–72.

Lima SL. 1995. Back to the basics of anti-predatory vigilance: the group-size effect. Animal Behaviour 49:11–20.

Offord C. 2018. AACR Apologizes for Delays in Retracting Papers. The Scientist.

Patience GS., Patience CA., Blais B., Bertrand F. 2017. Citation analysis of scientific categories. Heliyon 3.

Steen RG. 2011. Retractions in the scientific literature: is the incidence of research fraud increasing? Journal of medical ethics 37:249–53.

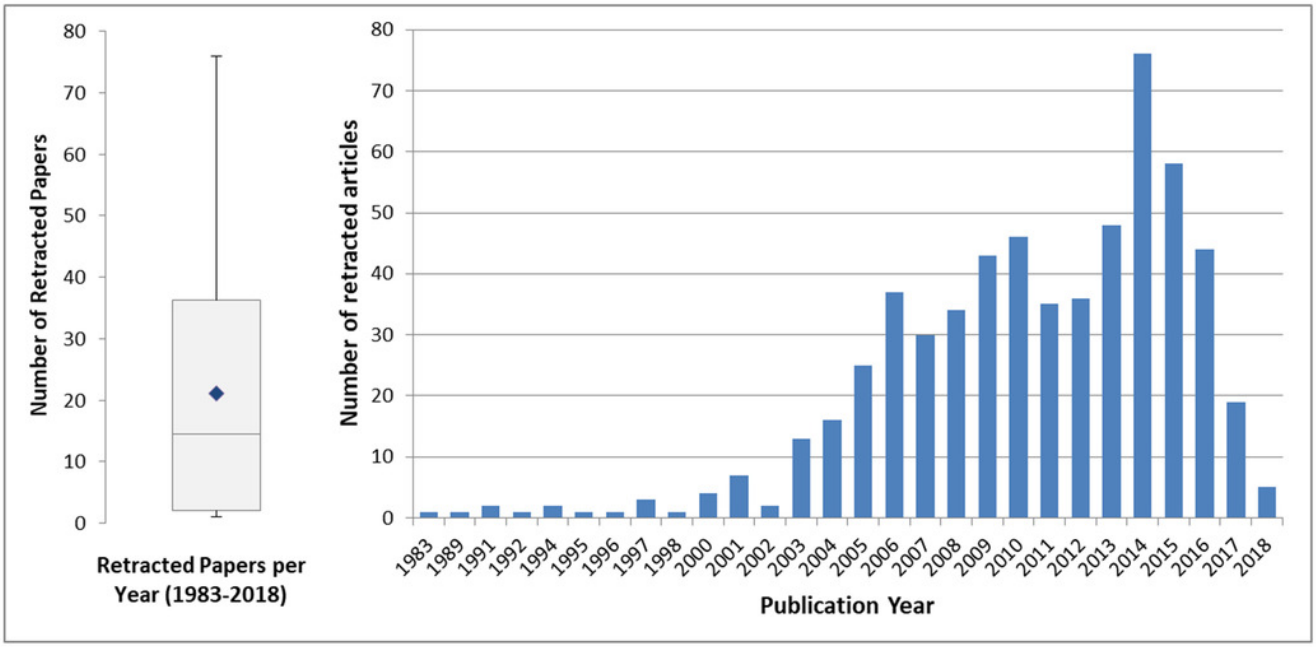
Steen RG., Casadevall A., Fang FC. 2013. Why has the number of scientific retractions increased? PloS one 8:e68397.

Ugarte A., Bouche G., Meheus L. 2014. Inconsistencies and questionable reliability of the publication “immunotherapy of metastatic colorectal cancer with vitamin D-binding protein-derived macrophages-activating, GcMAF” by Yamamoto et al. Cancer immunology, immunotherapy : CII 63:1347–8.

Wikipedia contributors. 2018.Linus’s Law --- {Wikipedia}. Available at https://en.wikipedia.org/w/index.php?title=Linus%27s_Law&oldid=860361874

Figure 1

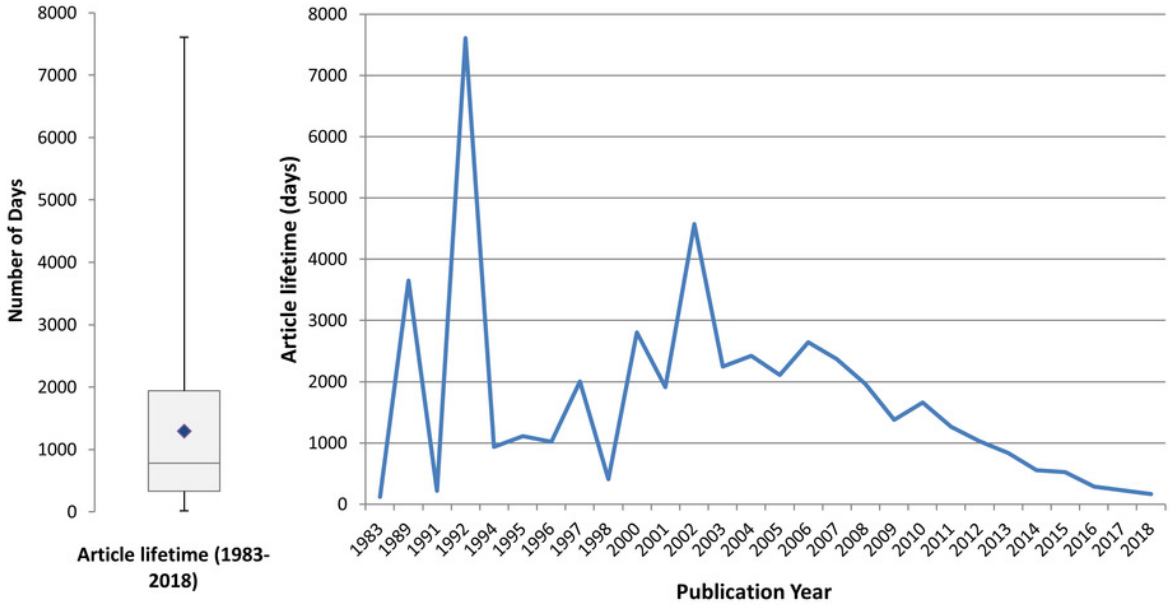
Annual distribution of retracted oncology articles in Pubmed 1983- 2018



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27427v1 | CC BY 4.0 Open Access | rec: 12 Dec 2018, publ: 12 Dec 2018

Figure 2

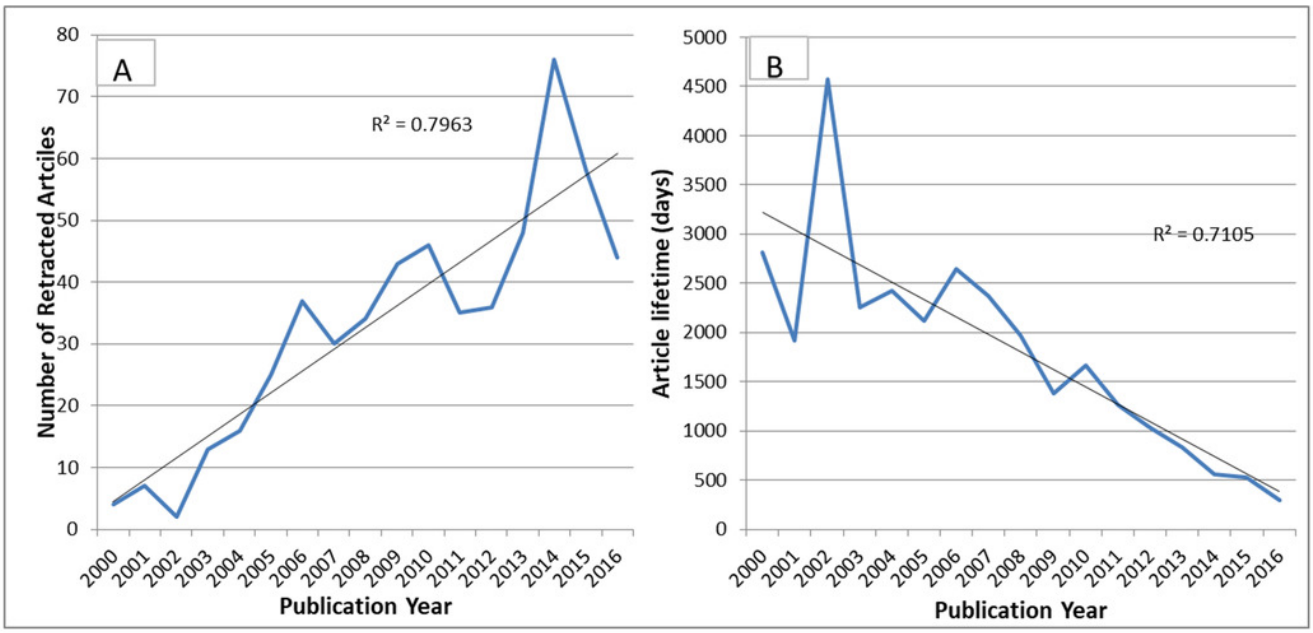
Mean article lifetime of retracted oncology articles in Pubmed 1983 - 2018



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27427v1 | CC BY 4.0 Open Access | rec: 12 Dec 2018, publ: 12 Dec 2018

Figure 3

Trends in retracted article counts and article lifetimes 2000 - 2016



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27427v1 | CC BY 4.0 Open Access | rec: 12 Dec 2018, publ: 12 Dec 2018

Figure 4

Publication and retraction rates

(A) Publication rate of all oncology articles in Pubmed (2000-2016). (B) Retracted articles per 10k oncology articles in Pubmed (2000-2016).

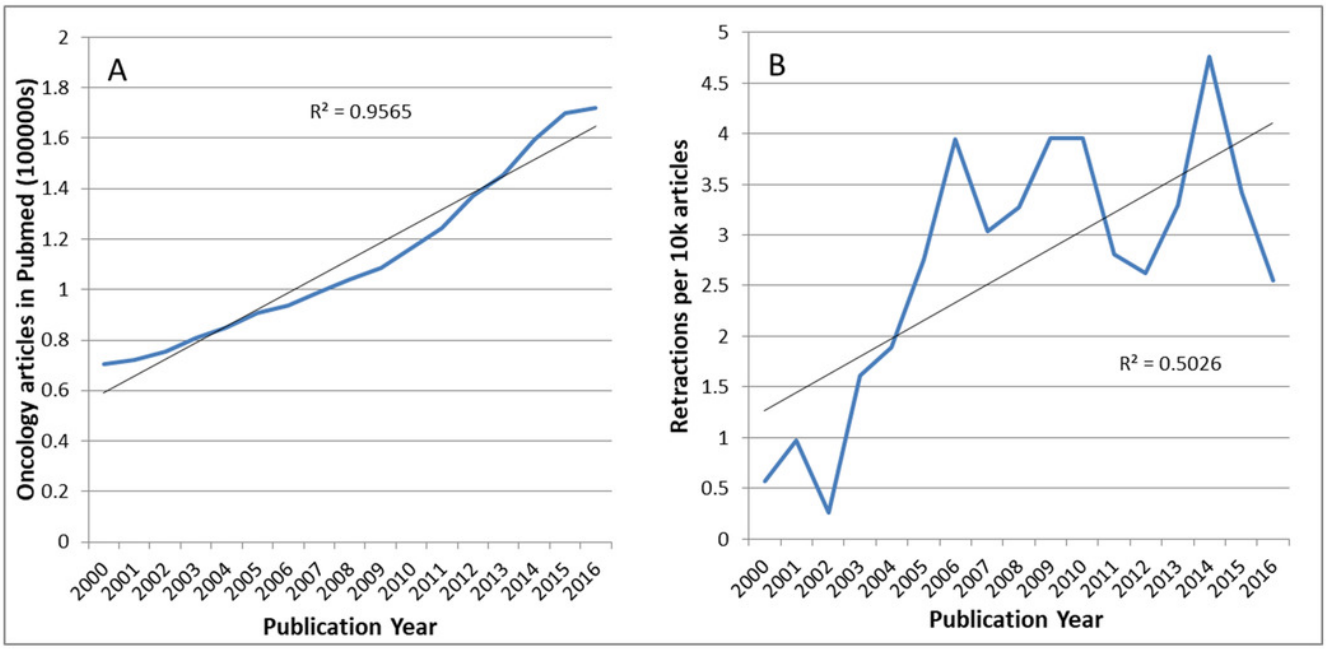


Figure 5

Authors and retractions

Number of authors per retracted oncology article in Pubmed (1983-2018). In red, the mean of number of authors per paper in the period 2000 - 2016.

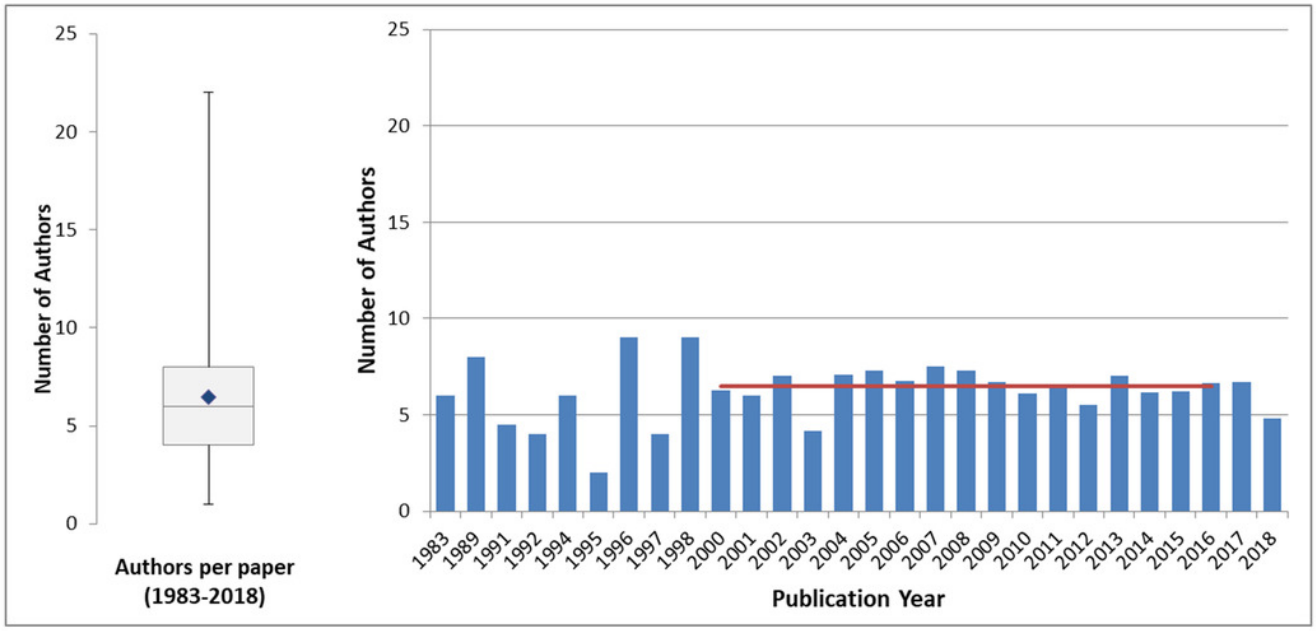


Figure 6

Distribution of authors/retractions

(A) Distribution of number of articles per author (1983 - 2018). (B) Distribution of author counts per article (1983 - 2018).

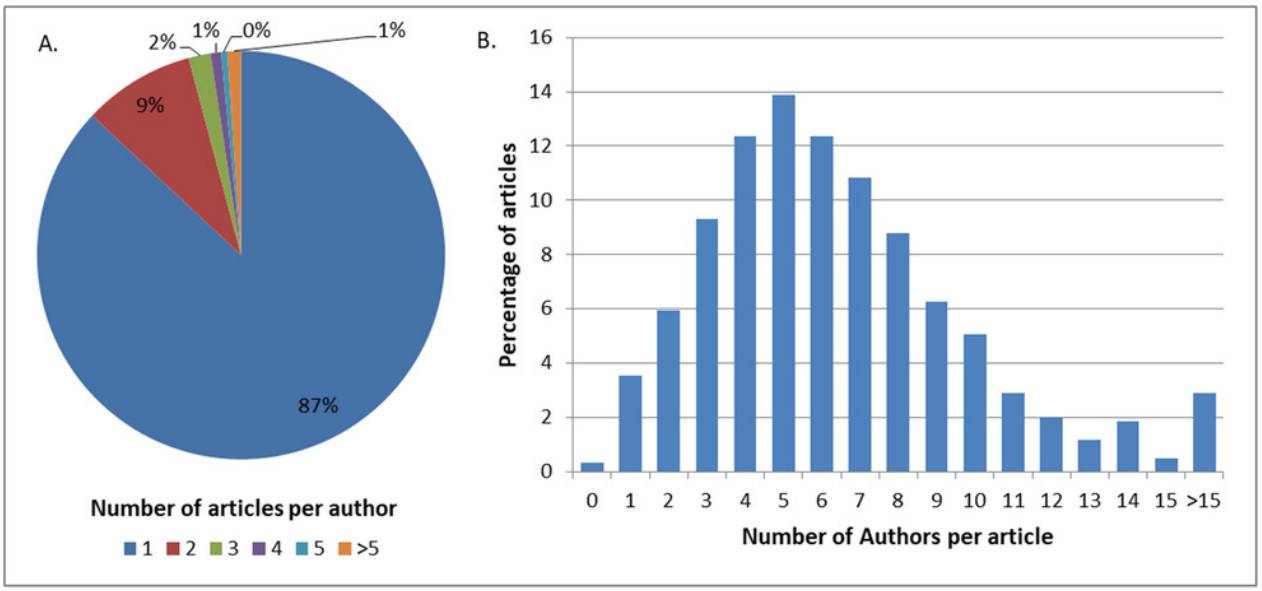
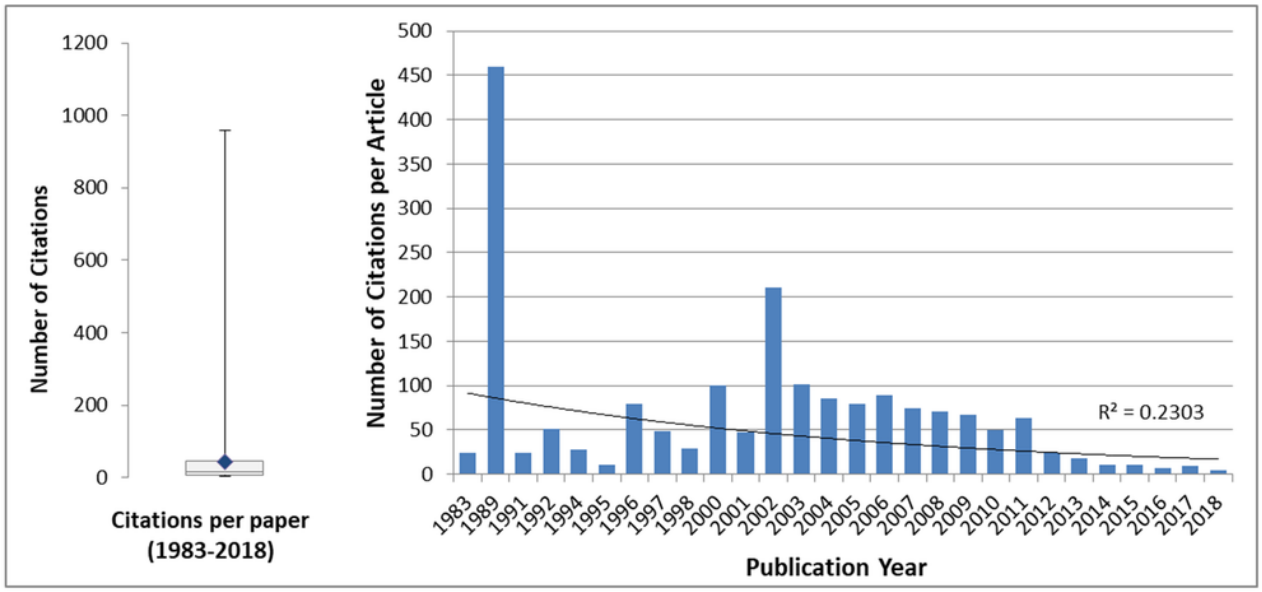


Figure 7

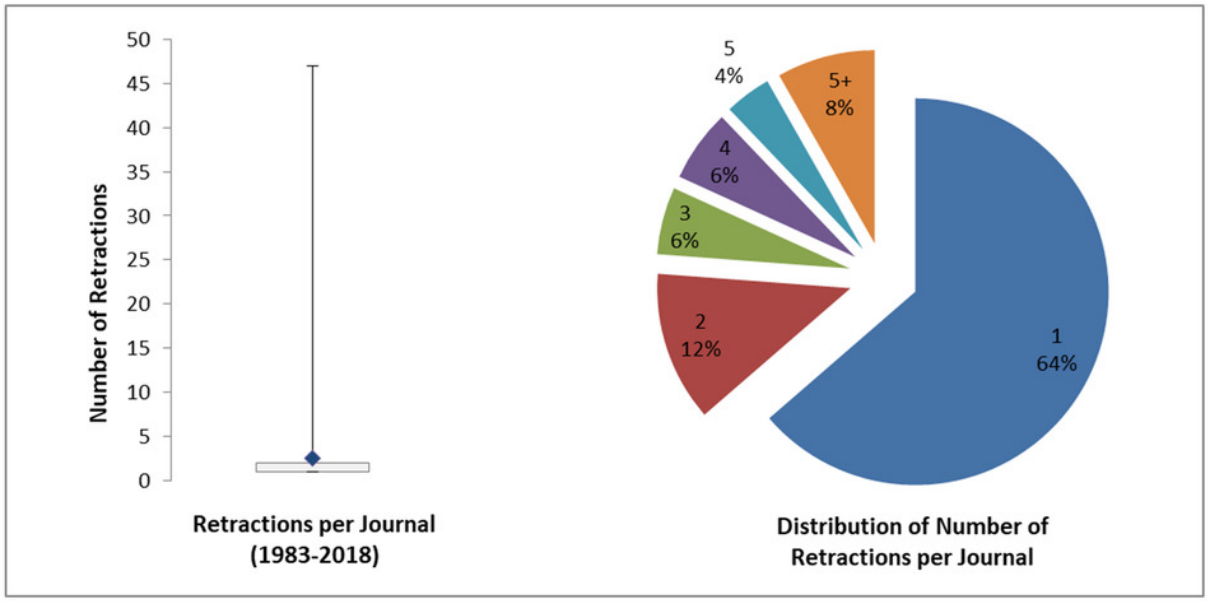
Citations per retracted oncology article in Pubmed (1983 - 2018)



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27427v1 | CC BY 4.0 Open Access | rec: 12 Dec 2018, publ: 12 Dec 2018

Figure 8

Distribution of Retractions by Journal, 1983 - 2018



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27427v1 | CC BY 4.0 Open Access | rec: 12 Dec 2018, publ: 12 Dec 2018