Accountability in Research
Policies and Quality Assurance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gacr20

For how long and with what relevance do genetics articles retracted due to research misconduct remain active in the scientific literature

Rafael Dal-Ré & Carmen Ayuso

To cite this article: Rafael Dal-Ré & Carmen Ayuso (2020): For how long and with what relevance do genetics articles retracted due to research misconduct remain active in the scientific literature, Accountability in Research, DOI: 10.1080/08989621.2020.1835479

To link to this article: https://doi.org/10.1080/08989621.2020.1835479

Published online: 30 Oct 2020.

ACCOUNTABILITY IN RESEARCH
https://doi.org/10.1080/08989621.2020.1835479

For how long and with what relevance do genetics articles retracted due to research misconduct remain active in the scientific literature

Rafael Dal-Ré[a] and Carmen Ayuso[b,c]

[a]Epidemiology Unit, Health Research Institute-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid, (IIS–FJD, UAM), Madrid, Spain; [b]Department of Genetics and Genomics, Health Research Institute-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid, (IIS–FJD, UAM), Madrid, Spain; [c]Center for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain

ABSTRACT

We aimed to quantify the number of pre- and post-retraction citations obtained by genetics articles retracted due to research misconduct. All retraction notices available in the Retraction Watch database for genetics articles published in 1970–2016 were assessed. The reasons for retraction were fabrication/falsification and plagiarism. The endpoints were the number of citations of retracted articles and when and how journals reported on retractions and whether this was published on PubMed.

Four hundred and sixty retracted genetics articles were cited 34,487 times; 7,945 (23%) were post-retraction citations. Median time to retraction and time to last citation were 3.2 and 3 years, respectively. Most (96%) had a PubMed retraction notice, One percent of these were totally removed from journal websites altogether, and 4% had no information available on either the online or PDF versions. Ninety percent of citations were from articles retracted due to falsification/fabrication. The percentage of post-retraction citations was significantly higher in the case of plagiarism (42%) than in the case of fabrication/falsification (21.5%) ($p<0.001$). Median time to retraction was shorter (1.3 years) in the case of plagiarism than for fabrication/falsification (4.8 years, $p<0.001$). The retraction was more frequently reported in the PDFs (70%) for the fabrication/falsification cases than for the plagiarism cases (43%, $p<0.001$). The highest rate of retracted papers due to falsification/fabrication was among authors in the USA, and the highest rate for plagiarism was in China.

Although most retractions were appropriately handled by journals, the gravest issue was that median time to retraction for articles retracted for falsification/fabrication was nearly 5 years, earning close to 6800 post-retraction citations. Journals should implement processes to speed-up the retraction process that will help to minimize post-retraction citations.

KEYWORDS
Research misconduct; genetics; citations; retraction notices; Retraction Watch database

CONTACT Rafael Dal-Ré rafael.dalre@quironsalud.es Unidad de Epidemiología, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz, Hospital Universitario, Universidad Autónoma de Madrid, 28040 Madrid, Spain

Supplemental data for this article can be accessed on the publisher's website.

Introduction

Article corrections and retractions are the two main pillars of the self-correcting approach of science. Data suggest that some 4 out of 10,000 articles are retracted (Brainard and You 2018) and that this rate is increasing in recent years (Bhatt 2020). Authors or editors retract articles due to ethical or methodological flaws in their execution and/or publication or because of other range of reasons, indicating that the scientific community in many cases must no longer rely on these text findings. One useful metric for measuring dissemination is the number of citations received by an article. However, it is well known that the retraction of an article could be followed by a rapid decline in the number of citations, although it could be cited as if the work were still valid (Fang, Steen, and Casadevall 2012). The latter could happen, on the one hand, if authors are not informed of the retraction, even though many journals and databases such as PubMed and Web of Science usually mark the titles of retracted articles. Nevertheless, retractions are not displayed clearly and consistently across platforms (Bakker and Riegelman 2018). On the other hand, authors could cite retracted articles when citing secondary sources—the "cut and paste" method (Madlock-Brown and Eichmann 2014)—leaving them unaware that they have been retracted. Yet, the delay between the publication of the article and that of the retraction notice could be considered the most relevant factor to ensure the correct approach to future research, and the use of only valid results for appraising and synthesizing evidence.

Research misconduct is commonly defined as fabrication, falsification, or plagiarism (OSTP 2000; ALLEA 2017). These three types of research misbehavior are considered especially serious because they distort the research record (ALLEA 2017). Other unethical research practices (e.g., mistreatment of research subjects, failure to disclose conflicts of interest, misrepresentation of data) have been added to these three by some national regulations (Resnik, Rasmussen, and Kissling 2015), research integrity codes and different stakeholders (Dal-Ré et al. 2020), but there is a lack of consensus on whether any additional examples of research misbehavior should be added to these three categories. Research misconduct has a personal, economic, social, and scientific impact (Lu et al. 2013; Stern et al. 2014; Mongeon and Lariviere 2016), jeopardizing future research and even patient care (Garmendia et al. 2019).

Genetics is relevant to all life-science disciplines, with thousands of investigators across many areas of specialization within this branch of science. To the best of our knowledge, only one previous study on retracted genetics articles describing in detailed the reasons and the times needed for retraction has been published (Dal-Ré and Ayuso 2019). One third of retracted genetics articles were retracted due to research misconduct (Dal-Ré and Ayuso 2019).

No reports have focused on citations received by genetics articles retracted due to research misconduct or how journals reported these retractions to their readership. To address this knowledge gap, an in-depth analysis of these two topics was deemed warranted, especially when in 2009, the Committee on Publication Ethics (COPE) issued guidelines to standardize the publication of retractions (Wager et al. 2009).

In this study, we aimed to assess the number of citations obtained by retracted genetics articles—both pre- and post-retraction—due to research misconduct (i.e., fabrication, falsification, plagiarism), and the types of retraction announcements published by journals and on PubMed. A comparison of these variables between articles retracted due to falsification/fabrication and articles retracted due to plagiarism was also conducted.

Materials and methods

This cross-sectional study was based on a search of the Retraction Watch database (RWdb) (Retraction Watch Database 2020) conducted on January 14–16, 2019 for articles published between 1 January 1970 and 31 December 2016. The following descriptors were applied: Nature of notice: "retraction"; article types: "case report," "clinical study," "commentary/editorial," "letter," "meta–analysis," "research article," and "review article," using the Boolean operator "OR." Subjects: "genetics." This strategy allowed us to exclude other types of articles such as "conference/abstract/paper," "dissertation/thesis," and "government publication." Then, as is predefined on the RWdb, the "reasons for retraction" searched for were as follows: falsification/fabrication or plagiarism (Supplemental material-1).

We retrieved the following information from the RWdb: PubMed ID or DOI of the original paper, date of publication of the original article and that of the retraction notice, journal, and country (ies) of origin of the authors. All Google Scholar citations for a given article were included in the study except those referring to patents. This meant that in addition to journal articles, we also retrieved other types of documents such as theses, books, and preprints. All citations written in English, French, Italian, Portuguese and Spanish were checked. Considering that a manuscript could be in the editorial process when a retraction notice was published, providing a 6-month "safety" period was a reasonable approach (Himmelstein 2016; Hamilton 2019). Hence, only articles published (first publication online or on paper) six or more months after the retraction notice was published were considered to be instances of post-retraction publication. All post-retraction publications were considered to be written after the authors would have ready access to the retraction notice. The publication date of each document was ascertained to be considered a post-retraction publication; those that could not be ascertained (e.g., owing to language issues) were not included as post-retraction publications. How journals informed readers at the article level that the text had been retracted was assessed in all retracted articles. This could be found on the journal's website in the online version of the article, and in the article's PDF (as a watermark—either transparent or opaque— on each article page or a note on the first page). A retracted article was classified as "removed" if the original webpage for the article and the retraction noticed were found, but the article itself had been removed; also, PDF versions could be removed (not available for download) but the online version of the article could be available.

In addition, we checked whether PubMed contained information related to articles that had been retracted through specific retraction notices. The Journal Citation Reports database was used to determine the journal impact factor. Since this information is only available for articles published from 1997 onwards, we noted the journals' 2017 impact factor (or the most recent impact factor available).

The search on Google Scholar was conducted between February 22 and 9 March 2019. To provide articles a minimum of 8 months to be cited after their retraction, and considering the above-mentioned "safety" period, only articles retracted up to 31 December 2017 were included in the analysis. Data were retrieved and checked by one author (RDR); the other author (CA) checked for data consistency using a random sample (by means of a random integer generator; https://www.random.org/) of 25% of all retracted articles and found no inconsistencies.

Statistical analyses

Categorical data were reported using frequencies and percentages. Continuous variables were described using medians and interquartile (Q1, Q3). Descriptive statistics of the number of citations (total number, and those occurring pre- and post-retraction) due to both types of research misconduct (falsification/fabrication, plagiarism), and time to analysis (i.e., time between the retraction date and the date of analysis), time to retraction (i.e., time between article publication and the retraction date), and time to last citation (i.e., time elapsed between the retraction date and the last citation) were calculated for the whole sample (genetics articles) and both types of research misconduct. Comparison of time to analysis, time to retraction, and time to last citation between retractions for falsification/fabrication and plagiarism were compared by means of the Mann–Whitney U test. We used the chi-square test to compare the pre- and post-retraction percentages of citations and the number of articles with at least 50% of citations 6 months after the publication date of the retraction notice that was retracted due to falsification/fabrication with those retracted due to plagiarism.

We used descriptive statistics to study the information available to readers via the online version, PDF version, and on PubMed for those genetics articles retracted on the grounds of the two types of research misconduct, as well as the authors based in the five countries with the highest number of retracted articles. To determine whether genetics articles retracted due to falsification/fabrication were reported to readers differently compared with retractions due to plagiarism, a comparison was conducted by means of the Fisher exact test (retraction reported in the article's online version) or chi-square test (retraction reported in the article's PDF version or on PubMed). The journal's impact factor for genetics articles retracted for falsification/fabrication and those retracted for plagiarism were compared using the Mann–Whitney U test.

A Bonferroni correction was applied for multiple testing. All tests were two-tailed; $p < 0.01$ was considered statistically significant. Statistical analyses were performed using R.3.5.1 (R Foundation for Statistical Computing, Vienna, Austria).

Results

The search yielded 460 retracted articles that were published in 237 different journals and received 34,487 citations (Table 1) (Supplemental material-2). The number of citations received before the retraction notices were published was 3.3 times higher than those received after retraction [26542 (77%) vs 7945 (23%)]. One out of four retracted articles received at least 50% of its citations after retraction. The median time to retraction was 3.2 years, while the median time to the article's last citation was 3 years. In 86% of cases, journals informed readers of the retraction of the articles' online full version, whereas this was the case in 62% of PDF versions (56% with a watermark, 6% with an accompanying note). Almost all retracted genetics articles (96%) had a PubMed retraction notice. Retracted articles were published in journals with a median impact factor of 4 (Table 1). There were 18 cases (4%) of retracted articles with no information available on either the articles' online version or the PDF. Most articles (81%, 376/460) were retracted after 2009.

Only 4 (0.9%) retracted articles had both falsification/fabrication and plagiarism issues; as a result, we had to consider 464 retracted articles when we split them into the two types of research misconduct. Of all retracted articles, 72% (332/464) were due to falsification/fabrication issues and 28% to plagiarism issues (Figure 1; Table 2). The 332 articles retracted due to falsification/fabrication were cited 31,515 times, whereas the 132 for retracted due to plagiarism were cited 3,368 times. Post-retraction citations were more frequent (42%) in the plagiarism cases than in the falsification/fabrication cases (21.5%; $p < 0.001$). Median time to retraction was statistically significantly lower ($p < 0.001$) with articles retracted due to plagiarism

Table 1. Genetics articles retracted due to research misconduct (up to March 2019). Variables regarding citations, how journals informed about the retraction, and JCR impact factor. N = 460.

Number of citations N (%)—median (Q1, Q3)	
Total	34487 – 29 (9, 75)
Before publication of the retraction notice (pre-retraction publication)	26542 (77.0) –18 (4, 56)
After publication of the retraction notice (post-retraction publication)[a]	7945 (23) – 6 (2, 17)
Time (years) to analysis[b]. Median (Q1, Q3)	4.7 (2.6, 8.1)
Time (years) to retraction data[c]. Median (Q1, Q3)	3.2 (1.3, 6.4)
Time (years) to last citation[d] Median (Q1, Q3)	3 (2, 6)
Year of publication of retracted articles Median (range)	2008 (1988 – 2016)
Year of publication of the retraction notices Median (range)	2014 (1990 – 2017)
Last year with citation (s) Median (range)	2018 (1998 – 2019)
Number of articles with ≥ 50% of post–retraction citations [e] N (%) [95% CI]	115 (25) [21.1, 29.2]
Information that the article has been retracted appearing in the online full text version N (%) [95% CI]	
Yes	396 (86.1) [82.9, 89.2]
No	59 (12.8) [9.7, 15.7]
Article removed	5 (1.1) [0.4, 2.5]
Information that the article has been retracted in the PDF N (%) [95% CI]	
Yes (watermark)	258 (56.1) [51.6, 60.6]
Yes (note)	28 (6.1) [3.9, 8.2]
No	49 (10.7) [7.8, 13.4]
Article removed	48 (10.4) [7.6, 13.1]
No checked (paywall access)	77 (16.7) [13.3, 20.2]
Retraction reported on PubMed N (%) [95% CI]	
Yes	441 (95.9) [94.1, 97.7]
No	19 (4.1) [2.3, 5.9]
Countries of origin [f] *N*	USA 191, China 93, Japan 47, India 30, UK 22
Journals' JCR impact factor [g] Median (Q1, Q3)	4.0 (2.9, 7.4)
Journals with high (>20) JCR impact factor : N	Science 13, Cell 12, Nature 7, Cell Metabolism 1, Lancet Oncology 1,

(a) articles published (first publication online or on paper) 6 or more months after publication of the retraction notice; (b) time to analysis: time from retraction date to the date of analysis; (c) time to retraction: time from article publication to retraction date; (d) time to last citation: time from retraction date to last citation; (e) Number of citations 6 months after publication of the retraction notice; (f) Only the top five countries are mentioned; (g) N = 450; JCR: Journal Citation Reports

(1.3 years) than for those retracted due to falsification/fabrication (4.8 years). Similarly, 46% of articles retracted due to plagiarism received at least 50% of post-retraction citations, whereas this happened with only 17% ($p < 0.001$) of articles retracted due to falsification/fabrication.

There were no differences in how journals reported retractions on articles' online versions between those linked to falsification/fabrication and those due to plagiarism. Conversely, in PDFs, journals reported statistically significantly ($p < 0.001$) more retractions when these were due to falsification/fabrication than when the retraction was due to plagiarism (Table 3). In fact, only 43% of plagiarized articles reported the retraction on their PDFs, whereas this happened in 70% of articles retracted on the grounds of falsification/fabrication.

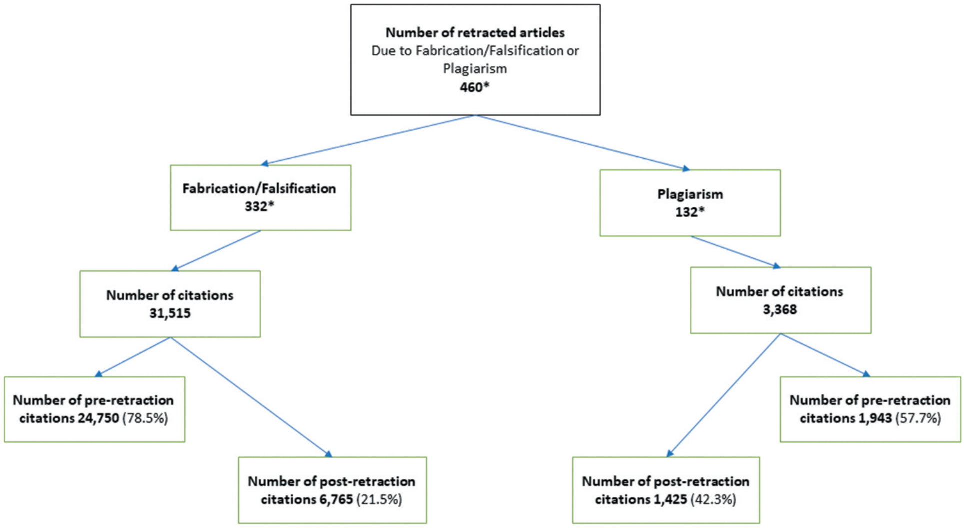

Figure 1. Flowchart. Genetics articles hosted on Retraction Watch database published in 1970–2016, retracted due to fabrication/falsification or plagiarism, and number of citations. *Of a total of 460 retracted articles, four of them were retracted for fabrication/falsification *and* plagiarism; that is why there are 332 retracted articles due to fabrication/falsification and 132 for plagiarism.

The percentage of articles without open access (paywalled access) was 2 times higher for articles retracted due to plagiarism (27%) than for falsification/fabrication (13%). Retraction notices indexed in PubMed were more frequent ($p < 0.001$) for cases of falsification/fabrication (98%) than for cases of plagiarism (91%). The USA was the country of origin of 53% of articles retracted due to falsification/fabrication, whereas most retractions due to plagiarism were from China (46%).

Finally, journals that had retracted genetics articles due to falsification/fabrication had a significantly higher ($p < 0.001$) impact factor (4.4 median; Q1, Q3: 3.8, 9.5; N = 329) than those that had retracted articles due to plagiarism (2.3 median; Q1, Q3: 1.4, 3.2; N = 121).

Discussion

This is the first study on citations received by retracted genetics articles and the first to conduct an in-depth comparison between the two types of research misconduct. Available data show that between 1996 and 2016 some 910,000 genetics articles were published (Scimago 2020). Our data showed that out of 460 genetics articles published between 1988 and 2016 that were eventually retracted up to 31 December 2017, falsification/fabrication was 2.5 times more

Table 2. Articles retracted* due to research misconduct. Number of citations** (total, before and after the publication of retraction notices), time from retraction to date of the analysis, and to last time the article was cited. Publication year of the retracted articles and year of the last citation (up to March 2019).

	Type of research misconduct	
Variable	Falsification/fabrication (N = 332)	Plagiarism (N = 132)
No. of citations: N (%) – median (Q1, Q3)		
Total	31515 (100) – 46 (19, 94.3)	3368 (100) – 7.5 (2, 16.3)
Before publication of the retraction notice (pre-retraction publication)	24750 (78.5)[b] – 32 (12, 74.3)	1943 (57.7)[b] – 3 (1, 8)
After publication of the retraction notice (post-retraction publication)[a]	6765 (21.5)[b] – 8 (2, 20)	1425 (42.3)[b] – 3 (0, 9.3)
Time (years) from retraction date to analysis. Median (Q1, Q3)	4.8 (2.7, 8.1)	4.5 (2.4, 7.7)
Time (years) from article publication to retraction date. Median (Q1, Q3)	4.8 (2, 7.5)[c]	1.3 (0.5, 2.7)[c]
Time (years) from retraction date to last citation (or time to last citation). Median (Q1, Q3)	4 (2, 6)	3 (1, 6)
Number of articles with ≥ 50% of post-retraction citations [a] N (%) [95% CI]	55 (16.6)[d] – [12.6–20.6]	61 (46.2)[d] – [37.7–54.7]
Year of publication of retracted articles: Median (range)	2008 (1988 – 2016)	2012 (1998 – 2016)
Year of publication of retraction notices: Median (range)	2014 (1990 – 2018)	2014 (2002 – 2017)
Lastyear with citation (s): Median (range)	2018 (1998 – 2019)	(1) (2001 – 2019)

*There were 460 retracted articles, but 4 of them were due to both falsification/fabrication and plagiarism. That is why the number of articles retracted for these two reasons totals 464. This is also the reason why in Table 2 the number of citations totals 34,883 (for 464 retracted articles), and not 34,487 (for 460 retracted articles) as shown in Table 1.

**21 retracted articles (6 for falsification/fabrication; 15 for plagiarism) received no citations at all.

(a) articles published (first publication online or on paper) 6 or more months after publication of the retraction notice; (b), (c), and (d): $p < 0.001$.

The reasons for retraction as per the Retraction Watch database nomenclature: Falsification/Fabrication = Falsification/fabrication of data, images, or results, and manipulation of results or images; Plagiarism = Euphemism for plagiarism and plagiarism of article, data, images, or text. See Supplemental data

Table 3. Articles retracted* due to research misconduct. Information available to readers on retracted articles on the online full-text version, PDF and on PubMed, and the five countries of origin with the largest number of retracted articles (as of March 2019).

	Types of research misconduct	
Variable	Falsification/fabrication (N = 332)	Plagiarism (N = 132)
Information that the article has been retracted in online full text version: N (%) [95% CI]		
Yes	291 (87.7) [84.1, 91.2]	109 (82.6) [76.1, 89]
No	37 (11.1) [7.8, 14.5]	22 (16.7) [10.3, 23]
Article removed	4 (1.2) [0.3, 3.1]	1[d] (0.8) [0, 4.1]
Information that the article has been retracted in the PDF: N (%) [95% CI]		
Yes (watermark)	208 (62.7) [57.4, 67.9][b]	53 (40.2) [31.8, 48.5][b]
Yes (note)	24 (7.2) [4.4, 10][b]	4 (3.0) [0.8, 7.6][b]
No	33 (9.9) [6.7, 13.2][b]	16 (12.1) [6.6, 17.7][b]
Article removed	24 (7.2) [4.4, 10][b]	24 (18.2) [11.6, 24.8][b]
Not checked (paywalled access)	43 (13) [9.3, 16.6][b]	35 (26.5) [19, 34][b]
Retraction reported on PubMed: N (%) [95% CI]		
Yes	325 (97.9) [96.3, 99.4][c]	120 (90.9) [86, 95.8][c]
No	7 (2.1) [0.6, 3.7][c]	12 (9.1) [4.2, 14][c]
Country of origin[a]: N	USA 176, Japan 48, China 33, UK 20, India 17	China 61, USA 15, India 13, Iran 12, South Korea 9

*There were 460 retracted articles, but 4 of them were due to both falsification/fabrication and plagiarism. That is why the number of articles retracted for these two reasons total 464.

(a) Many articles had authors from more than one country; only the top five countries are mentioned; (b) and (c) $p < 0.001$; (d) The journal was not found. The reasons for retraction as per the Retraction Watch database nomenclature: Falsification/Fabrication = Falsification/fabrication of data, images or results, and manipulation of results or images; Plagiarism = Euphemism for plagiarism and plagiarism of article, data, images or text. See S1.

frequently the reason for retraction than plagiarism. This figure falls between 1.1 for retracted psychology papers (Stricker and Günther 2019) and 4.4 for retracted biomedical articles (Fang, Steen, and Casadevall 2012).

Our data showed that the median number of citations per retracted genetics article, 29, was significantly associated with the reason for retraction: 46 for falsification/fabrication and 7.5 for plagiarism. The statistically significantly different time to retraction of articles on the grounds of falsification/fabrication versus plagiarism seemed to play a role in the latter finding (see below). Other studies on retractions for all possible reasons have found medians of 4, 6, 8, and 18 citations per retracted article, in chemistry and material science (Coudert 2019), orthopedics (Yan et al. 2016), obstetrics and gynecology (Chambers, Michener, and Falcone 2019), and oncology (Bozzo et al. 2017), respectively. These differences could be due to the specific features of the discipline under study, the different data sources (e.g., Google Scholar, PubMed, Web of Science), the distribution of reasons for retraction in the sample under evaluation—which, as shown here, could be critical—and the median time elapsed from the retraction to the date of analysis—which was 4.7 years in this study. This latter is seldom reported in other studies. These three variables should be taken into consideration by future investigators when conducting retraction analyses on any discipline.

The COPE guidelines state that retraction notices "should be published promptly in order to minimize harmful effects from misleading publications" (Wager et al. 2009). Our data showed that the median time from article publication to retraction was 3.2 years—with a maximum of 32 years—but with statistically significant differences between those retractions due to falsification/fabrication (4.8 years) and plagiarism (1.3 years). This could be explained, at least in part, by the pervasive use of software products that check for plagiarism. Fang , Steen, and Casadevall (2012) also found that the retraction of articles due to plagiarism took almost half of the time needed for retractions due to falsification/fabrication. Other studies have shown that the median time to retraction varied between 1 and 2.5 years (Faggion et al. 2018; King et al. 2018; Chambers, Michener, and Falcone 2019; Coudert 2019; Mena et al. 2019; Kardes et al. 2020; Rapani et al. 2020), with maximum delays of 13 (Mena et al. 2019), 16 (King et al. 2018), 18 (Coudert 2019) and 21 (Chambers, Michener, and Falcone 2019) years. Most (81%) retractions were between 2010 and 2016—after the COPE guidelines had been published—which translated to an increase in the number of retractions in the last years of the analysis. Similarly, recent studies on retractions in different disciplines found that between 54% and 80% (Yan et al. 2016; Bozzo et al. 2017; Wang et al. 2017; Faggion et al. 2018; Chambers, Michener, and Falcone 2019; Coudert 2019; Mena et al. 2019; Kardes et al. 2020) of all retractions were in the last 5–10 years of the analysis.

Journals should not remove retracted articles (Wager et al, 2009; ICMJE, 2019) and retraction notices should be linked to the article, in both the PDF and online version (Wager et al. 2009). Our data show that only 1% of retracted genetics articles were totally removed from journals' websites. This is a remarkably lower percentage than that found in other studies, which varied from 8% (Yan et al. 2016; Bakker and Riegelman 2018) to 9% (Bozzo et al. 2017). However, an additional 4% did not provide any information whatsoever—neither in the online version nor in the PDF— that the article had been retracted. In our study, journals informed about retraction in the online full version in 86% of cases. With regards to stamping a watermark on the pages of the PDF to indicate retraction of the article, this was the case in 56% of all retracted genetics articles with available PDFs, that is at the lower limit of data reported for other disciplines between 56% and 89% (Yan et al. 2016; Bozzo et al. 2017; Wang et al. 2017; Bakker and Riegelman 2018; Mena et al. 2019; Nair et al. 2020; Rapani et al. 2020).

The overall rate of post-retraction citations of retracted genetics articles, 23%, is quite different from the 38% observed in urology (Mena et al. 2019)—the only one article that, to the best of our knowledge, reported this figure. It should be highlighted that the percentages of pre-/post-retraction citations in plagiarized and subsequently retracted genetic articles differed from those articles retracted on the basis of falsification/fabrication. Due, in part, to this fact, 46% of plagiarized genetics papers received at least 50% of post-retraction citations—as opposed to 17% of papers retracted due to falsification/fabrication—suggesting that many authors citing such retracted papers were unaware of the publication of the retraction notices or simply disregarded them. This might also be related to another finding of this study, specifically that journals handling articles retracted due to plagiarism informed readers with significantly less transparency than those handling retractions due to falsification/fabrication. Thus, authors who directly accessed the PDFs of retracted articles would most likely have been misguided in more instances when consulting articles retracted due to plagiarism than those consulting papers retracted due to falsification/fabrication. However, this would not have been the case for those authors accessing retracted articles through the journals' websites, since they would have had access to the online full-text version, which reported similarly on retracted articles regardless of the reason for retraction.

PubMed indexed retraction notices for 96% of retracted genetics articles, although this was statistically significantly more common when reporting on articles retracted due to falsification/fabrication (98%) than to plagiarism (91%). This is a very high percentage when compared to an analysis of retracted mental-health articles retrieved from the RWdb, with only 62% having retraction notices on PubMed (Bakker and Riegelman 2018). The retracted genetics articles were published in journals with a median impact factor of 4; other studies have reported median impact factors of retracted articles of 2.4–3.6 (King et al. 2018; Chambers, Michener, and Falcone 2019; Mena et al. 2019; Nair et al. 2020). Of note, the median impact factor for genetics articles retracted as a result of falsification/fabrication was almost double (4.4) that of genetics articles retracted as a result of plagiarism (2.3). Most retracted articles were by authors from the USA and China, as reported elsewhere (Fang, Steen, and Casadevall 2012; Wang et al. 2017; Chambers, Michener, and Falcone 2019; Coudert 2019; Mena et al. 2019).

Limitations

This study has several limitations. First, we used only one source of retracted articles, the RWdb. Although this is the most comprehensive and largest database of retracted articles (Brainard and You 2018), we could have missed some retracted genetics articles. However, there are data suggesting that the RWdb hosted more retracted articles than PubMed and the Web of Science (Bakker and Riegelman 2018). Second, the RWdb did not distinguish between fabrication and falsification as reasons for retraction. However, the seven platforms (databases and aggregators, such as EBSCO, PubMed, Web of Science, and Scopus) assessed by Bakker and Riegelman (2018) inconsistently met the COPE guidelines on standardization of retraction notices. Therefore, this limitation was offset by the fact that using the information on retraction notices would have posed another issue since it is not standardized, varies substantially across journals, and could even be misleading through ambiguous wording (King et al. 2018; Chambers, Michener, and Falcone 2019). Third, we used only Google Scholar to search for citations from the retracted papers, when there are other available databases. Nevertheless, it has been shown that Google Scholar is much more consistent in finding citations than the Web of Science and Scopus across all disciplines (Hamilton 2019; Martin-Martin et al. 2018). In addition, it would have been altogether inefficient to check the 34,487 Google Scholar citations (or a meaningful percentage) against the tens of thousands from other databases to find out which ones could be added. However, we admit that some citations have been missed. Fourth, citations could be positive, neutral, or negative, and we did not check this. Checking this feature in close to 34,500 articles was inefficient and beyond the scope of this study. In any case, only a low percentage of citations are negative—i.e., referring to the retracted article as such. Some 2% (Catalini, Lacetera, and Oettl 2015) of citations are negative, although this figure could be higher in specific disciplines [e.g., 7% in radiation oncology (Hamilton 2019) or 10% in dentistry (Rapani et al. 2020)]. Fifth, we assessed the relevant data of those publications that appeared in Google Scholar in English and four Romance languages (i.e., French, Italian, Portuguese, and Spanish) and in those written in other languages that provided the data of interest. Finally, this study, as in the case for all studies conducted based on retractions, did not include other flawed papers that have not been retracted or even articles whose authors have been found guilty of research misconduct but with no retraction notice published (Drimer-Batca, Iaccarino, and Fine 2019).

It is of concern that many retracted genetics articles—as shown in this study—and articles retracted in other disciplines (Yan et al. 2016; Bozzo et al. 2017; Wang et al. 2017; Bakker and Riegelman 2018; Faggion et al. 2018; Stricker and Günther 2019; Chambers, Michener, and Falcone 2019; Coudert 2019; Mena et al. 2019) inconsistently followed the COPE guidelines. The fact that a significantly lower percentage of post-retraction citations were linked to falsification/fabrication (22%) than to plagiarism (42%) should be highlighted since the former—which indicates misrepresentation of the truth (Kuroki 2018)—is a much more serious type of research misconduct than the latter—which implies a betrayal of trust (Kuroki 2018)—in that it misleads future research, wastes human and technical resources, and potentially leading to harm for patients (Steen 2011; Garmendia et al. 2019). Both, however, are impacting the trust of society in science.

Several proposals have been put forward to minimize the number of citations of retracted articles. A number of stakeholders should be involved in this. One could think that peer reviewers, as experts in the field, could check the references of manuscripts and flag to the editor any retracted article mentioned in them. However, it seems likely that most reviewers will decline to do this, as they do not tend to crosscheck the information of the manuscript with that of the registry to prevent outcome reporting bias (Mathieu, Chan, and Ravaud 2013). Editorial teams could screen manuscript references before publication to remove them or to determine whether a given reference was retracted (Bar-Ilan and Halevi 2017): this would be a useful preventive measure to minimize the consequences of post–retraction citation. This task –which is now facilitated by the RWdb– takes a similar approach to that of the detection of image manipulations in the editorial process that a number of journals have already put in place. In addition, editors should work to improve the consistency and transparency of journals with regards to article retractions (Chambers, Michener, and Falcone 2019). From the publishers' side, they could accept that content aggregators make retraction notices readily available across proprietary database platforms (Zietman, Yom, and Braverman 2019). Finally, it would be of great help for journals if the COPE were to issue guidelines on what editors should do when realizing that a manuscript has included post-retraction citations (Bar-Ilan and Halevi 2017). In any case, in the current scenario, the number of post-retraction citations could be minimized if authors check all the references mentioned in their manuscripts and if journals follow the COPE guidelines (Wager et al. 2009).

Conclusion

This work has shown that close to 25% of citations of genetics articles retracted due to research misconduct occurred after retraction of the article. However, the most important issue was the long time elapsed between the publication of the article and its retraction for articles retracted for falsification/fabrication, which implies betrayal of truth. Thus, the median time to retraction for articles retracted for this reason was nearly 5 years; in this period, these retracted articles earned close to 6,800 post-retraction citations. These figures for articles retracted for plagiarism were remarkably different: time to retraction was more than one year, with more than 1,400 post-retraction citations. Journals should implement processes to enhance transparency, standardize retractions, and minimize post-retraction citations.

Acknowledgments

We thank Ignacio Mahillo PhD (Health Research Institute-Fundación Jiménez Díaz, University Hospital, Universidad Autónoma de Madrid, Madrid, Spain) for conducting the statistical analyses. We thank the University Chair UAM-IIS-FJD of Genomic Medicine for funding the open access publication charges of this paper, and Oliver Shaw for the English review of this article.

Data availability

All data supporting this research are provided in the article and as supplemental material.

Disclosure statement

No potential conflicts of interest was reported by the authors.

Funding

This study required no funding.

References

ALLEA–All European Academies. 2017. *The European Code of Conduct for Research Integrity*. Revised ed. Berlin. Accessed 27 June 2020. https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

Bakker, C., and A. Riegelman. 2018. "Retracted Publications in Mental Health Literature: Discovery across Bibliographic Platforms." *Journal of Librarianship and Scholarly Communication* 6: eP2199. doi:10.7710/2162-3309.2199.

Bar-Ilan, J., and G. Halevi. 2017. "Post Retraction Citations in Context: A Case Study." *Scientometrics* 113 (1): 547–565. doi:10.1007/s11192-017-2242-0.

Bhatt, B. 2020. "A Multi-perspective Analysis of Retractions in Life Sciences." *bioRxiv*. June 1. doi:10.1101/2020.04.29.063016.

Bozzo, A., K. Bali, N. Evaniew, and M. Ghert. 2017. "Retractions in Cancer Research: A Systematic Survey." *Research Integrity and Peer Review* 2 (1): 5. doi:10.1186/s41073-017-0031-1.

Brainard, J., and J. You. 2018. "Rethinking Retractions." *Science* 362 (6413): 391–393. doi:10.1126/science.362.6413.390.

Catalini, C., N. Lacetera, and A. Oettl. 2015. "The Incidence and Role of Negative Citations in Science." *Proceedings of the National Academy of Sciences USA* 112 (45): 13823–13826. doi:10.1073/pnas.1502280112.

Chambers, L. M., C. M. Michener, and T. Falcone. 2019. "Plagiarism and Data Falsification are the Most Common Reasons for Retracted Publications in Obstetrics and Gynaecology." *BJOG* 126 (9): 1134–1140. doi:10.1111/1471-0528.15689.

Coudert, F. X. 2019. "Correcting the Scientific Record: Retraction Practices in Chemistry and Material Sciences." *Chemistry of Materials* 31: 3593–3598.

Dal-Ré, R., and C. Ayuso. 2019. "Reasons and Time to Retraction of Genetics Articles (1970–2018)." *Journal of Medical Genetics* 56 (11): 734–740. doi:10.1136/jmedgenet-2019-106137.

Dal-Ré, R., L. M. Bouter, P. Cuijpers, C. Gluud, and S. Holm. 2020. "Should Research Misconduct Be Criminalized?" *Research Ethics* 16 (1–2): 1–12. doi:10.1177/1747016119898400.

Drimer-Batca, D., J. M. Iaccarino, and A. Fine. 2019. "Status of Retraction Notices for Biomedical Publications Associated with Research Misconduct." *Research Ethics* 15 (2): 1–5. doi:10.1177/1747016118820496.

Faggion, C. M., R. S. Ware, N. Bakas, and J. Wasak. 2018. "An Analysis of Retractions of Dental Publications." *Journal of Dentistry* 79: 19–23. doi:10.1016/j.jdent.2018.09.002.

Fang, F. C., R. G. Steen, and A. Casadevall. 2012. "Misconduct Accounts for the Majority of Retracted Scientific Publications." *Proceedings of the National Academy of Sciences U S A* 109 (42): 17028–17033. doi:10.1073/pnas.1212247109.

Garmendia, C. A., L. Nassar Gorra, A. L. Rodriguez, M. J. Trepka, E. Veledar, and P. Madhivanan. 2019. "Evaluation of the Inclusion of Studies Identified by the FDA as Having Falsified Data in the Results of Meta-analyses: The Example of the Apixaban Trials." *JAMA Internal Medicine* 179 (4): 582—4. doi:10.1001/jamainternmed.2018.7661.

Hamilton, D. G. 2019. "Continued Citation of Retracted Radiation Oncology Literature-Do We Have a Problem?" *International Journal of Radiation Oncology Biology Physics* 103 (5): 1036–12. doi:10.1016/j.ijrobp.2018.11.014.

Himmelstein, D. 2016. "The History of Publishing Delays." *Satochi Village*. [Blog post]. Accessed 27 June 2020. https://blog.dhimmel.com/history-of-delays/

ICMJE,2019. "Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. December. http://www.icmje.org/recommendations/

Kardes, S., W. Levack, K. Özkuk, E. A. Aydin, and S. S. Karabulut. 2020. "Retractions in Rehabilitation and Sport Sciences Journals: A Systematic Review." *Archives of Physical Medicine and Rehabilitation*. May 10. Online Ahead of Print doi:10.1016/j.apmr.2020.03.010.

King, E. G., I. Oransky, T. E. Sachs, A. Farber, D. B. Flynn, A. Abritis, J. A. Kalish, and J. J. Siracuse. 2018. “Analysis of Retracted Articles in the Surgical Literature.” *American Journal of Surgery* 216: 851–855.

Kuroki, T. 2018. “New Classification of Research Misconduct from the Viewpoint of Truth, Trust and Risk.” *Accountability in Research* 25 (7–8): 404–8. doi:10.1080/08989621.2018.1548283.

Lu, S. F., G. Z. Jin, B. Uzzi, and B. Jones. 2013. “The Retraction Penalty: Evidence from the Web of Science.” *Scientific Reports* 3 (1): 3146. doi:10.1038/srep03146.

Madlock-Brown, C. R., and D. Eichmann. 2014. “The (Lack Of) Impact of Retraction on Citation Networks.” *Science and Engineering Ethics* 21 (1): 127–137. doi:10.1007/s11948-014-9532-1.

Martin-Martin, A., E. Orduna-Malea, M. Thewall, and E. Delgado Lopez-Cozar. 2018. “Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories.” *Journal of Informetrics* 12 (4): 1160–1177. doi:10.1016/j.joi.2018.09.002.

Mathieu, S., A. W. Chan, and P. Ravaud. 2013. “Use of Trial Register Information during the Peer Review Process.” *PLoS One* 356 (4): e59910. doi:10.1371/journal.pone.0059910.

Mena, J. D., M. Ndoye, A. J. Cohen, P. Kamal, and B. N. Breyer. 2019. “The Landscape of Urological Retractions: The Prevalence of Reported Research Misconduct.” *BJU International* 124 (1): 174–179. doi:10.1111/bju.14706.

Mongeon, P., and V. Lariviere. 2016. “Costly Collaborations: The Impact of Scientific Fraud on Co–authors’ Careers.” *Journal of the Association for Information Science and Technology* 67 (3): 535–542. doi:10.1002/asi.23421.

Nair, S., C. Yean, J. Yoo, J. Ledd, E. Delphin, and D. C. Adams. 2020. “Reasons for Article Retraction in Anesthesiology: A Comprehensive Analysis.” *Canadian Journal of Anesthesia* 67 (1): 57–63. doi:10.1007/s12630-019-01508-3.

OSTP. 2000. “Federal Policy on Research Misconduct.” *Federal Register* 65: 76260–76264. Accessed 27 June 2020. https://www.govinfo.gov/content/pkg/FR-2000-12-06/pdf/00-30852.pdf

Rapani, A., T. Lombardi, F. Berton, V. Del Lupo, R. Di Lenarda, and C. Stacchi. 2020. “Retracted Publications and Their Citation in Dental Literature: A Systematic Review.” *Clinical and Experimental Dental Research* 6 (4): 383–90. doi:10.1002/cre2.292.

Resnik, D. B., L. M. Rasmussen, and G. E. Kissling. 2015. “An International Study of Research Misconduct Policies.” *Accountability in Research* 22: 249–66.

Retraction Watch Database, version: 1.0.5.5. Accessed 27 June 2020. http://retractiondatabase.org/RetractionSearch.aspx?&AspxAutoDetectCookieSupport=1

Scimago Journal and country RANK. Accessed 27 June 2020. https://www.scimagojr.com/countryrank.php?category=1311

Steen, R. G. 2011. “Retractions in the Medical Literature: How Many Patients are Put at Risk by Flawed Research?” *Journal of Medical Ethics* 37 (11): 688–692. doi:10.1136/jme.2011.043133.

Stern, A. M., A. Casadevall, R. G. Steen, and F. C. Fang. 2014. “Financial Costs and Personal Consequences of Research Misconduct Resulting in Retracted Publications.” *eLife* 3: e02956. doi:10.7554/eLife.02956.

Stricker, J., and A. Günther. 2019. “Scientific Misconduct in Psychology. A Systematic Review of Prevalence Estimates and New Empirical Data.” *Zeitshrift für Psychologie* 227: 53–63.

Wager, E., V. Barbour, S. Yentis, and S. Kleinert. 2009. “Guidance from the Committee on Publications Ethics (COPE).” *Croatian Medical Journal* 50: 532–535. Also available at COPE Website. Accessed 27 June 2020. http://www.publicationethics.org/files/retraction%20guidelines.pdf

Wang, J., J. C. Ku, N. M. Alotaibi, and J. T. Rutka. 2017. "Retraction of Neurosurgical Publications: A Systematic Review." *World Neurosurgery* 103: 809–814.e1.

Yan, J., A. MacDonald, L. P. Baisi, N. Evaniew, M. Bhandari, and M. Ghert. 2016. "Retractions in Orthopaedic Research: A Systematic Review." *Bone and Joint Research* 5 (6): 263–268. doi:10.1302/2046-3758.56.BJR-2016-0047.

Zietman, A. L., S. S. Yom, and L. C. Braverman. 2019. "Making Sure Retractions Matter." *International Journal of Radiation Oncology Biology Physics* 103 (5): 1043—4. doi:10.1016/j.ijrobp.2019.01.001.