Propagation of Error: Citations to Problematic Research*

Ken Cor Gaurav Sood

February 26, 2022

*We are grateful to Danielle Portnoy, Xiaoran Huang, and Suriyan Laohaprapanon for assisting us with research, and to Andrew Gelman, Kabir Khanna, and Daniel Stone for offering valuable comments. The data and scripts for replicating the analysis are posted at: https://github.com/recite/propagation_of_error

Summary. To shed light on how often scientists base their claims on problematic research, we exploit data on cases where problems with research are broadly publicized. Using data from over 3,000 retracted articles and over 74,000 citations to these articles, we find that at least 31% of the citations to retracted articles happen a year after they have been retracted and that about 91% of the post-retraction citations note no concern with the cited article. Our results have implications for the design of scholarship discovery systems and scientific practice more generally.

Keywords. Citation Behavior, Retractions, Scientific Integrity, Scientific Misconduct, Scientometrics

1 Introduction

Citations are the bedrock of the scientific process. Scientists use citations to give credit for being first (“x, y, and z have studied a”), to debate methods and inferences (“the method used in study x fails to account for s”), as evidence (“x shows a”, “we use data from x for our meta-analysis”), and to contextualize results (“our results are consistent with results from y”). And unless the researcher notes problems with cited research, citations cue that the data, results, inferences, or in some cases, the entire article can be trusted.

When researchers cite articles with serious errors without acknowledging the errors, problems ensue. When erroneous research is cited as evidence (e.g., Chang et al. 2013; Torsvik et al. 2010), it cues that the evidence for the claim is good. Such citations unduly increase the reader’s confidence in the result or argument. In the extreme, a reader may become persuaded that the point being buffeted by a citation to problematic research is right. The reader, generally another academic, may go on to write other articles influenced by the incorrect point, citing the erroneous article for support, or may share the point as fact with colleagues, students, and practitioners, propagating the error.

For example, consider an article published in April 2005 by Rubio et al. (2005) reporting a disturbing finding in *Cancer Research*. They report discovering that stem cells can spontaneously transform into cancerous cells during *in vitro* experiments. The finding was a blow to research in the use of stem cells to treat cancer. By 2010, according to *Web of Science*, the article had been cited over 375 times.

In August 2010, the article was retracted (De la Fuente et al. 2010). The authors had been unable to replicate the result, and there was mounting evidence that transformations like the one reported were due to a basic error: cross-contamination during cell culturing. While this episode can be seen in a favorable light, given that the errors were caught and a retraction notice was issued, the story does not end there.

As of January 2022, the article had been cited another 400 plus times, with most citations noting no concern with the original work. For instance, a year after being retracted, Firinci et al. (2011) published an article in *International Immunopharmacology* citing Rubio et al. as basis for warning scientists that stem cells can spontaneously transform. Two years after the retraction, Kosaka et al. (2012) published an article in *Cancer Gene Therapy* in which they cited the evidence from Rubio et al. as a hurdle to implementation of the treatment they found to be effective. Three years after the retraction, Chang et al. (2013) published an article in *Aesthetic Plastic Surgery* citing Rubio et al. to argue that spontaneous transformation of human stem cells remains a risk. Finally, as recently as December 2021, Marofi et al. (2021) cited the article as an example of research that is needed to investigate transformation and tumor formation. More generally, citations years after retraction are less rare than expected (Schneider et al. 2020; Van Der Vet and Nijveen 2016).

Citing erroneous research without acknowledging errors is problematic in other ways. Such citations give full credit to research (and researchers) when, at best, partial credit is deserved. And since citation tallies cue credibility, such citations make erroneous research appear more credible.

Lastly, sometimes research with serious errors is cited to acknowledge the source of the data. For instance, Lin et al. (2013) used data from "two retracted studies ... without acknowledgment of their retractions, both of which were for fraudulent data..." (p. 1, Paul et al. 2015) in a meta-analysis. In such cases, the consequence is obvious and extreme—the key findings in the published work are incorrect.

In this paper, we study how common citations to problematic research that do not note concerns with the original work are.

2 Citations to Problematic Research

Citing problematic research without noting the problems is an urgent issue in light of two empirical findings. First, the number of retractions—research where the problems are serious enough to warrant a retraction—is rising, even after controlling for the number of articles published (Grieneisen and Zhang 2012; Mena et al. 2019; Steen 2011; Wadhwa et al. 2021).[1] Second, the main reason research is retracted—major error and fraud—make post retraction citation a serious threat to the credibility of the scientific process (Bozzo et al. 2017; da Silva and Bornemann-Cimenti 2017; Grieneisen and Zhang 2012; Lee et al. 2021; Singh et al. 2014).

Because of these reasons, researchers have started to look at how citation rates change post retraction (Bar-Ilan and Halevi 2018; Rubbo et al. 2019; Panahi and Soleimanpour 2021; Wadhwa et al. 2021) and whether post retraction citations acknowledge problems in the retracted work (e.g., Bar-Ilan and Halevi 2017; Hamilton 2019; Heibi and Peroni 2021; Luwel et al. 2018). Bar-Ilan and Halevi (2018), for instance, use a sample of 994 retracted articles from `ScienceDirect` from a single month—October 2014—and find that the total number of citations to retracted articles grew over time. Similarly, Rubbo et al. (2019) use a sample of 238 retracted engineering articles and their citations and report the number of times the articles are cited pre and post retraction and find no apparent difference. Neither article notes whether citations after retraction note problems with the cited article. Some studies, however, consider whether post-retraction citations acknowledge problems. Bar-Ilan and Halevi (2017), for instance, find that, of the 283 citations to retracted case-study articles, a majority did not note problems with the article they cited. Similarly, Hamilton (2019) report that of 358 citations to retracted articles in the radiation oncology field, 92% referenced the research as legitimate. Finally, Luwel et al. (2018) studied pre- and post- retraction citations to a set of highly publicized retractions and found that over 95% of citations were positive or neutral in how the researchers characterized the cited work. Older

[1] Steen et al. (2013) posits that the rise in retractions is a result of declining quality of research and better ability to detect problems.

research has found similar results. A study using a database of 235 retracted biomedical articles found that nearly 94% of the citations after retraction treated research as valid (Budd et al. 1998) and work done as far back as 1990, exploiting a dataset of 82 retracted articles, came to similar conclusions (Pfeifer and Snodgrass 1990).

While studies of post-retraction citations consistently show high rates of positive or neutral references to retracted research following official notification, they all use small samples or samples that span a single discipline. The use of small, selective samples means that we cannot confidently say how widespread the problem is. With the exception of some studies like Luwel et al. (2018), analyses of the nature of citations to retracted articles also just focus on citations after retraction, which means that we cannot say how common citations that fail to acknowledge concerns are before the problems are publicized and if the rate changes after publicity. Finally, while researchers such as Rubbo et al. (2019) have conducted treatment versus control comparisons using non-parametric analyses of total citation counts, there has yet to be an analysis of whether citation count trends pre and post retraction differ across treatment and control group articles with a large dataset using a difference-in-differences design.

In all, there is growing evidence that retracted research continues to be cited in a way that does not acknowledge problems long after being retracted. Most of the evidence comes from studies that use small, selective samples. Moreover, the studies present data on post-retraction citations rather than the change in the number of citations over time and the quality of citations after a retraction. In all, there is sparse research on how the quality and quantity of citations change over time after retraction.

3 Methods

3.1 Hypotheses and Research Design

We expect the publication of the retraction notice to raise awareness about the retracted article. And we expect greater awareness to dramatically reduce citations to problematic research, especially citations that fail to acknowledge the problem. Formally, we hypothesize that retracted articles will receive no approving citations after retraction. We test this hypothesis by tallying the proportion of citations to retracted articles that are received post-retraction and use a two-tailed t-test to compare it against the hypothesis that the proportion is 0.

We are also interested in estimating how the citation rate for retracted articles changes after retraction. To estimate the impact of retraction on citation rates, we implement a difference-in-differences design. We compare citation rates of retracted articles before and after retraction to before and after citation rates of other articles that were published at the same journal at about the same time. We expect a sharp decline in citations to retracted articles post-retraction. And we expect other articles that were published in the same journal and at the same time as the retracted article to continue to be cited at roughly the same rates after the publication of retraction.

3.2 Data

Our dataset starts with over 3,000 retracted articles and nearly 74,000 citations to the retracted articles extracted from the Web of Science (WoS) database (Clarivate 2021a) up to August 2016.

We used WoS to assemble the set of 3,000 retracted articles and over 74,000 citations because it is one of the most comprehensive curated databases of academic research from all disciplines. WoS indexes articles from over 21,000 natural and social science journals and books and conference proceedings (Clarivate 2021b). WoS indexes over 82 million records, 126,000 books and 226,000 conferences (Clarivate 2021b). WoS contains key citation indices including the *Science Citation Index Expanded* (1900–present), *Social Sciences Citation Index* (1900–present), *Arts &*

Humanities Citation Index (1975–present), *Conference Proceedings Citation Index* (1990-present), *Book Citation Index* (2005–present), among others. For a full list of titles included in the *Science Citation Index Expanded*, *Social Sciences Citation Index*, *Arts & Humanities Citation Index*, and *Conference Proceedings Citation Index*, and a synopsis of the *Book Citation Index*. Most importantly for our work and unlike freely available databases like Google Scholar, WoS offers a way to filter results based on correction flags as well as the ability to download standardized article records containing extensive metadata.

To build the database of retracted articles, we started by creating a list of retraction notices. To do that, in August 2016, we searched WoS for titles containing the phrase "retraction of." The search yielded more than 14,000 records dating back to 1900. Using the "corrections" filter in WoS—it is a WoS flag for retraction and correction notices—we filtered the list to 4,085 retraction notices.[2]

Next, we wrote software to automatically search the WoS database for retracted articles using the information in the retraction notice records. Retraction notice records did not contain consistent titles to allow a simple search, but 99% of the retraction notices contained the year the original article was published, and 96% listed the authors of the original work. We used these two pieces of information along with the name of the publication to search the WoS for the original articles. The search resulted in a list of 3,776 articles. We could not locate the remaining 309 retracted articles.

Due to the variability in the information contained in the retraction notice records, the automated search process returned the wrong article in some cases. Our aim was to have zero false positives, even at the risk of some false negatives. With that aim, we created rules to flag

[2]Given the original data was collected in 2016, we do not have data from 2016 to 2021, a period during which we expect, based on the trends we show later in the paper, substantial growth in the number of retractions. As with any period analysis (albeit a long period), we cannot confidently say how our results will generalize into the present. When we look at the core problems that we conjecture underlie the pattern, we see many of them are still in place. And combined with the fact that the rate of retracted articles is sharply rising, it may well be the fact that things are substantially worse today.

potential false positives. First, if the list of authors of the retracted article did not match the list of authors for the relevant retraction notice record, we flagged the record as a potential false positive. Second, if the title of the retracted article did not contain the words “retracted” or “retraction,” we flagged it as a potential false positive.[3] Third, we parsed the title of the retracted notices to extract the title of the original retracted article, and we flagged articles where the titles did not match as potential false positives. We then reviewed the potential false positives, filtering out all records where we could not verify the match. This resulted in a set of 3,084 articles. Finally, we checked for duplicates. We found 55. This left us with 3,029 articles that served as our final retracted article sample.[4]

To get a list of citations to these articles, we used the WoS functionality that allows users to access the list of citations to articles. We wrote software to download citation records for each of the retracted articles automatically.

We also wrote software to create a control group sample using the 3,029 retracted article records. We used the publication title, date of publication, and research area fields to search for all publications that were published in the same journal at the same time in the same field as the retracted article. We then filtered the results using the WoS document type filter to research articles. From the resulting list of articles, we selected two records at random to serve as control group articles for the retracted article. We then looked up and downloaded the citations to the resulting list of control group articles using the WoS citation functionality. We were able to identify 5,755 control group articles for 2,952 of the retracted articles. These articles produced 201,862 citations up to August, 2016.

[3]Our data suggests that it has become common practice to revise the titles of original articles to indicate that they have been retracted. However, adherence may vary across disciplines. As a result, our sample may be biased in favor of disciplines where the adherence is greater.

[4]As an additional robustness check, we manually checked a random sample of 100 retracted articles to confirm that the article had indeed been retracted. We found that all of them were.

4 Results

Table 1 presents some descriptive statistics about the data. The first column presents data on retracted articles, and the second column presents data on articles that cite the retracted articles. On average, retracted articles were cited about 31 times in total. The average impact factor of the journals in which they were published is 6, and by 2016, it had been about nine years on average since they were published. The articles citing the retracted articles were published in less prestigious journals, with an average impact factor of about 5. And expectedly, the articles citing the retracted articles are a bit newer—by 2016, it had been about seven years since they were published, on average.

To describe the fields in which retractions occur, we augmented the Web of Science research field categorization scheme to classify the articles. There is one caveat. Sometimes papers cover more than one topic. We chose the first topic in these cases taking it to be the primary topic.

As Table 1 shows, 65% of the retracted articles were published in the Life Sciences and Biomedicine field. A distant second at 13% is the Physical Sciences, followed by Technology at 10.7% and the Social Sciences at 5.5%. One reason why a large majority of the retractions are from the Life Sciences and Biomedicine field may be simply because the field has more publications, but we cannot say anything definitely except that this is consistent with other studies that have examined the distribution of retractions across different research fields. The other clear (and expected) pattern in the data is that the field split of the citing articles is broadly the same as that of retracted articles.

Over the last thirty or so years, the number of retractions has increased sharply (see Figure 1). The first retraction notice that we have in our database is from 1989. That year and the decade after it, the number of retraction notices being published per year never crossed 20. Since then, there has been a sharp and accelerating rise in the number of retraction notices per year.

Variable	Retracted	Citing
Avg. Number of Citations	30.65	–
Avg. Journal Impact Factor	6.12	5.26
Avg. Number of Years Since Published	8.77	7.07
Field		
Arts & Humanities	.43%	.01%
Life Sciences & Biomedicine	65.21%	75.11%
Multidisciplinary	5.19%	7.40%
Physical Sciences	12.98%	10.62%
Social Sciences	5.45%	1.60%
Technology	10.74%	5.28%

Table 1: Summary Statistics of Retracted Articles and Articles Citing Retracted Articles

Between 2001, when 16 retraction notices were published, and 2015, last year for which we have complete data, there was a near 30 fold increase; a total of 451 retraction notices were published in 2015. The pattern that we find is consistent with results from Steen et al. (2013), who also find a rapid increase in retractions over time.

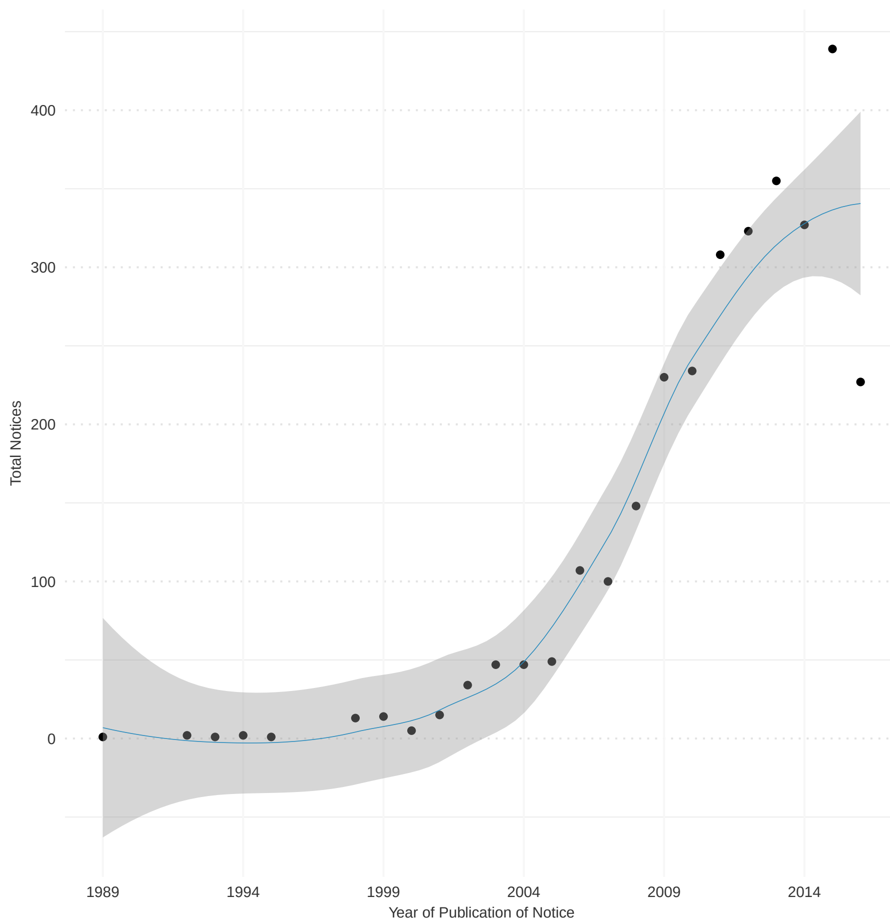

Figure 1: Retraction Notices Per Year

The rapid rise in the number of retractions shown above is likely the result of a combination of increasing production and improvements in detection. In all, there is an ever faster growing number of articles that should not have been published in the first place.

To understand why the articles are retracted, we coded a random sample of 100 retraction notices. 39% of the notices mentioned plagiarism as one of the major reasons for retracting the article. (Plagiarism includes self-plagiarism, duplication of data, words, and publishing the same or similar article in multiple journals.) Major errors or fraud contributed to another 51% of the retractions, with fraud alone accounting for 24% of the retractions. Ethics violations (2) and conflict over authorship or approval from other authors (5) contributed to the rest. The percentage of retractions attributable to major errors or fraud in our data is similar to the number obtained by other research on reasons for retraction in other corpora. For instance, a study of 1,112 Biomedicine articles retracted between 1997 and 2009 found that 55% were retracted for some type of misconduct (Budd et al. 2011) (see also Steen (2011)). In all, most articles are retracted because the research cannot be trusted.

These flawed articles often accrue a fair number of citations. In our data, the articles had been cited 39,792 times before being retracted, with another 34,000 citations and counting after retraction. (As hypothesized, the proportion of citations post-retraction is significantly different from zero.) This is not unsurprising, given that it took, on average, 2.85 years for the article to be retracted. The median time before the article was retracted was two years (see Figure 2), with 28.1% of the articles taking four or more years to be retracted. These numbers are similar to those obtained elsewhere (e.g., Steen et al. 2013).

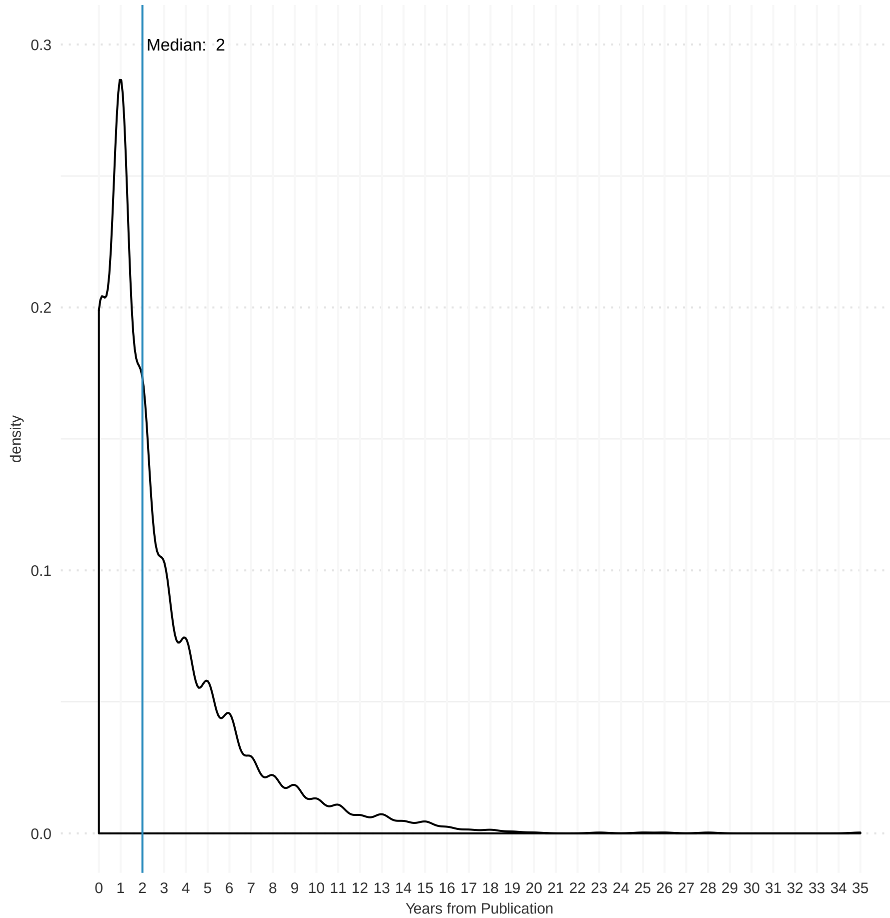

Figure 2: Time to Retraction

To investigate a secondary question of whether the greater readership of more prominent journals would mean that problematic articles are flagged more quickly, we estimated the relationship between the Journal Impact Factor (JIF) and average time to retraction. As Figure 3 shows, the relationship is flat—flawed articles in low-ranked journals are retracted about as quickly as flawed articles in higher-ranked journals.

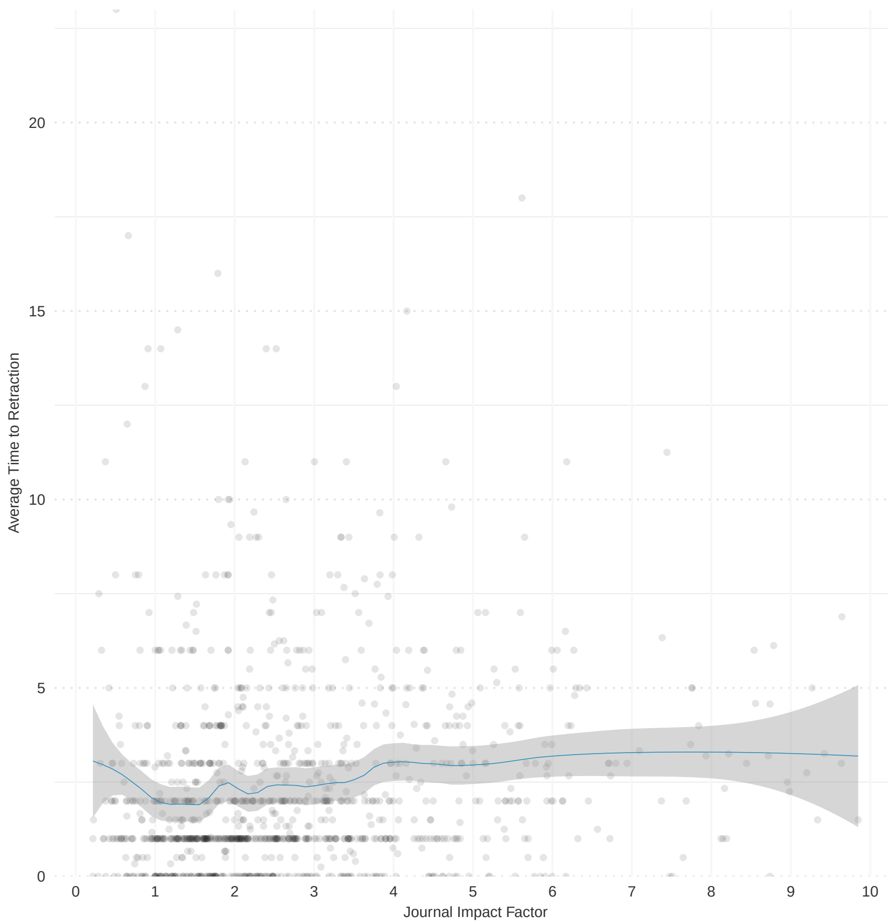

Figure 3: Relationship Between Journal Impact Factor and Time to Retraction

Given that most retracted articles are retracted because of serious error or fraud, we expect retracted articles to *never* be cited without acknowledgment of the problem a year or more—taking account of long publication windows—after the retraction notice has been published. However, retracted articles were cited another 22,932 times between the year after they were retracted and August 2016.[5] Thus, on average, the retracted articles received an additional 7.5 citations. Given the skew in retraction notices, with the bulk being published in recent years, these totals include very little post-retraction data for many of the articles. In other words, the results are a *lower bound* of the percentage of citations that likely happen after an article has been retracted.

[5]SI 1 presents some robustness checks based on how we used data on our database to code what is pre- and post-retraction. We manually coded 300 articles to estimate measurement error and found that error plausibly leads to a small (.5%) *reduction* in the proportion of citations we code as post-publication.

4.1 Estimating The Impact of Retractions on Citations

To explore the frequency of citations before and after retraction for the retracted articles and the control group articles, we plotted line graphs of the median number of citations per article per year against year from the publication of retraction notice (see Figure 4). We limit ourselves to 10 years before and after the publication of the retraction notice as the data are very sparse beyond that. For the retracted article sample, median citations per year decline when the retraction notice is published from 4 to 2 between the year before the retraction notice is published and the year after the retraction notice is published, followed by a plateauing. For the control group articles sample, the median citation count stays at 3 for the commensurate period.

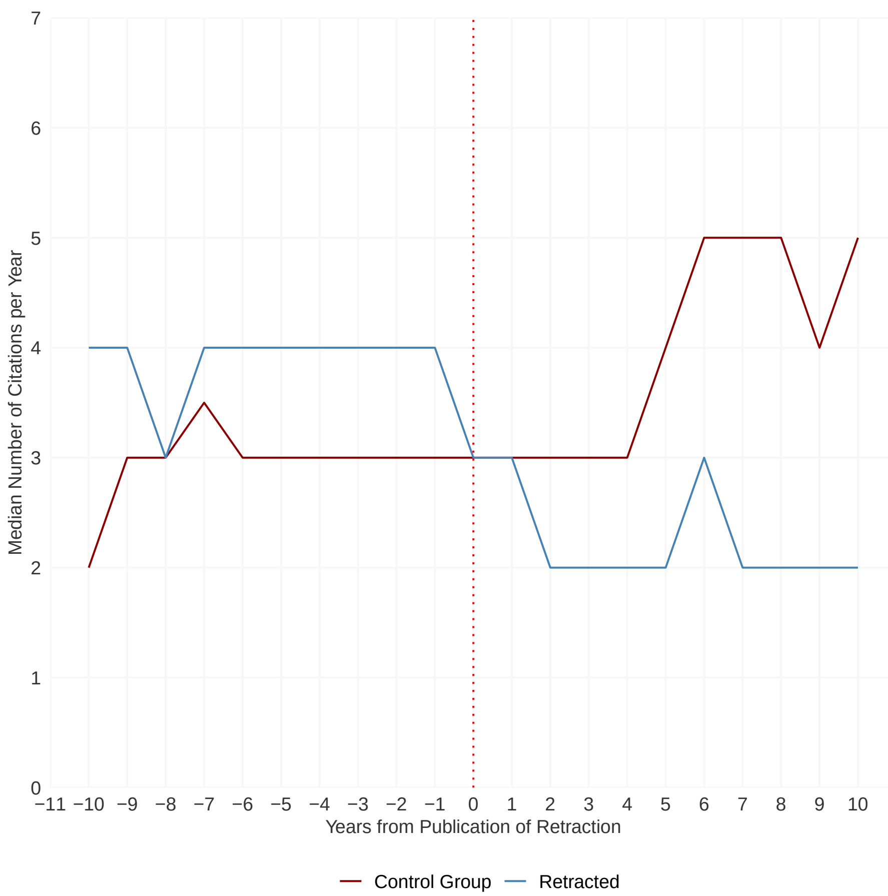

Figure 4: Comparison Between Median Number of Citations per Year to Control Group and Retracted Articles

Our main analysis is a Difference-in-Differences (DID) analysis. DID gives us a way to control for any confounding over time trends. We estimated whether the difference in citation rates of retracted and the paired unretracted articles changed after the retraction. (The implied treatment date for control group articles is the date when the paired retracted article is retracted.) To account for non-linearity in citations, we fit a Poisson model. In particular, letting i index articles and j index years, we regressed log of citations per year ($y_i j$) on whether or not the article is retracted (s_i) (1 = retracted), a dummy for years before or after the publication of the retraction notice (n_i) (1 = post-retraction), an interaction between the two (which captures the change in the difference between citations per year between retracted and control group articles), and an error term (u_j). To account for the dependence between observations induced by the fact that we observe the same article multiple times, we include random effects for each article that we take as normally distributed with 0 mean and $sigma^2$ variance. In all, we estimated the following model:

$$y_{ij} \sim Pois(\alpha + \beta_1 n_i + \beta_2 s_i + \beta_3 (n_i * s_i) + u_i)$$

$$u_i \sim N(0, \sigma^2)$$

The results of the model can be seen in Table 2. As Table 2 shows, on average a control group article is cited about 3.3 times per year before treatment. Before being retracted, retracted articles are cited slightly more (about 17% more). Post-treatment, the retracted articles receive nearly 50% fewer citations, but the citations to the control group articles remain roughly flat. The substantial decline in citations to retracted articles is reassuring, but we would like to emphasize that the post-retraction citation rate for retracted articles is still non-zero, with about a third of citations to retracted articles happening after retraction.

To estimate how many of the citations after retraction are *fail to acknowledge problems*, we

Table 2: Difference-in-Difference Analysis of the Impact of Publication of Retraction on the Number of Times Retracted vs. Unretracted Article is Cited

	Dependent variable:
	Citations per year
Treatment Date	-0.04^{***}
	(0.01)
Retracted or Not	0.2^{***}
	(0.03)
Retracted or Not*Treatment Date	-0.7^{***}
	(0.01)
Constant	1.2^{***}
	(0.02)
Observations	22,626
Akaike Inf. Crit.	48,156.5
Bayesian Inf. Crit.	48,196.7
Note:	*p<0.1; **p<0.05; ***p<0.01

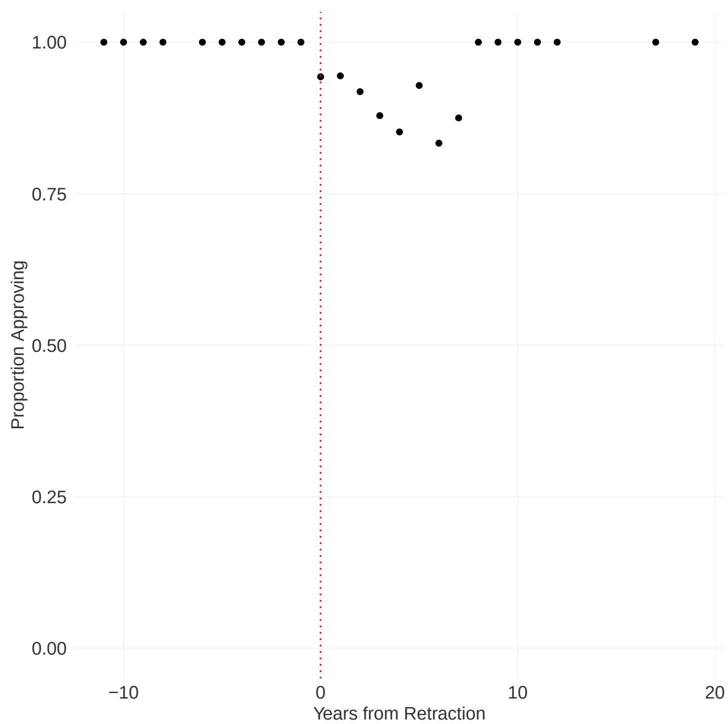

Figure 5: Proportion of Citations That Fail to Acknowledge Problems Per Year

coded a random sample of 100 articles that cited a retracted article pre-retraction and 275 articles that cited retracted research a year or more after the publication of the retraction notice. We could not locate 32 articles, leaving us with 343 total articles. There were no false positives. Of the 87 articles citing retracted articles before or in the same year the retraction notice was published, 85 (97.7%) did not acknowledge any problems with the research being cited. Of the 256 articles citing retracted articles the year after the retraction notice was published, 234 (91.4%) were did not mention problems - a slight trend downward. (Figure 5 plots the proportion of citations failing to acknowledge problems by year.) In all, the data suggest that retracted articles continue to be cited at non-zero rates after retraction, and they are still very much failing to acknowledge concerns.

5 Discussion

Only some of the points in a scientific paper rely on original work. For most points, scientists rely on work done by others. For instance, scientists rely on other research to buffet arguments, situate work in a research tradition, credit others for original arguments, data, and results, etc. (Erikson and Erlandson 2014). However, sometimes, the work done by others is problematic. For instance, sometimes, the claims in the cited work are not only wrong but made with fraudulent data. And we expect scientists to either not refer to such work or acknowledge the issues fully, especially after the errors have been publicized.

In this paper, we use a large dataset of retracted articles and their citations and a novel dataset of articles with a statistical error and their citations and discover that the citation rate to problematic research after publication of problems is lower but still significant.

There are a few reasons why these types of citations exist. The first set of reasons has to do with systems and infrastructure. For example, da Silva and Bornemann-Cimenti (2017) conjecture that many researchers may be unaware that articles have been retracted because databases and publishers do not often provide clear indications of retraction in a timely fashion. Static reference databases likely exacerbate the problem. Even when publisher databases mark papers as retracted, many researchers rely on static reference databases stored on their computers for citations. These databases sometimes contain articles that have since been retracted. For instance, Davis (2012) finds that personal Mendeley libraries contained 1,340 retracted articles. Journal publishers also do not seem to screen citations to retracted research in submitted manuscripts. Instead, they assume that scientists will cite appropriately and that the peer review process will screen the remaining problems.

The second category of possible reasons is more researcher-centered. For instance, there is a general bias in the scientific community to trust published work. This trust is likely buffeted by the belief that scientific misconduct is limited to a few bad people, which in turn may be driven by the fact that media reporting often focuses on personalities rather than processes. For example, cases of Diederik Stapel, who fabricated data behind at least 30 papers (Levelt et al. 2012), John Darsee, who faked data behind nearly 100 publications (Stewart and Feder 1987; Anderson et al. 2013; Wallis 1983), and Jan Hendrik Schön, who during a period in 2001 published a research paper every eight days based on fabricated data (Service 2003; Anderson et al. 2013) were heavily covered in the media. So were the cases of Andrew Wakefield, who published an article linking MMR vaccine to autism using fabricated data (Wakefield et al. 1998; Deer 2011; Godlee et al. 2011), and recently Michael Lacour, who published a paper in *Science* based on fabricated data (Broockman et al. 2015; McNutt 2015).[6] Each of these cases was framed as an example of misconduct by a bad actor, with the subtext being that a bad actor is an outlier.

Misconduct, however, is not limited to a few bad actors. A large anonymous survey of early- and mid-career scientists found that about 2% of scientists admitted to engaging in fabricating, falsifying, or plagiarizing in the last *three* years Martinson et al. (2005) (see also Titus et al. (2008)). Another study found that nearly 34% of the respondents in past surveys admitted to engaging in questionable research practices Fanelli (2009).

The other likely reason for trust in peer-reviewed research is that the rate of official retractions is extremely low. For instance, in a study of the nearly 9.4 million articles published between 1950 and 2004 and available on PubMed, only 596 had been retracted (Cokol et al. 2007). In all likelihood, however, the actual rate of serious errors in manuscripts is manifolds that rate. For instance, Cokol et al. (2007) estimate the rate at which articles ought to be retracted to be anywhere between 16.7 times to 167.8 times the actual rate. And these estimates do not account for research that involves harder-to-prove malpractice such as filing away non-significant results (Franco et al. 2014), conducting specification searches, and other more fundamental concerns like low power, which reduces the likelihood that a nominally statistically significant finding reflects

[6]Other prominent cases include that of William Summerlin, who painted mice rather than transplant skin (Basu 2006; Anderson et al. 2013), Woo Suk Hwang, who claimed to have cloned embryos, Eric Poehlman, who fabricated data behind at least ten papers and numerous grant applications.

a true effect (Button et al. 2013; Ioannidis 2005).

It is also likely that scientists cite erroneous research due to a lack of time stemming from the pressure to publish. Some scientists may also buckle to incentives to treat published research generously, given harsh (but accurate) judgments may provoke some reviewers. More generally, peer reviews tend to focus on instances where the authors fail to cite someone or miss the journal's formatting requirements more than concerns over incorrectly citing bad research. This means that there is little incentive to cite carefully.

These 'miscitations' are avoidable. Improving how the information about problematic research is generated and communicated can ameliorate the problem. For example, our study has resulted in the creation of a relatively large database of retracted research. Other groups are compiling even larger databases (e.g. RetractionWatch (2021)). These databases can be leveraged by publishers to screen submitted manuscripts for citations to retracted research. They could also be used to develop web browser add-ons for scholarly search engines such as Google Scholar and JSTOR to highlight problematic articles listed on a web page. In addition, tools that automatically create pull requests to personal bibliography libraries posted on open publication platforms like GitHub could be built. Taking steps like these should lead to significant improvements in reducing citations of problematic research.

To conclude, this study contributes to a rich literature on scholarly citation behavior. For example, Rekdal (2014) introduces the concept of *academic urban legends*— rumors that appear frequently and in complex and colorful ways that can usually be traced back to poorly employed sources—like the story of a decimal point error resulting in the belief that spinach is a good source of iron. Similarly, Lance et al. (2006) explores the basis for *methodological urban legends* in the form of four commonly employed criteria for the goodness of fit indices, reliability, inter-rater agreement indices, and eigenvalues. They find that much of the hype about these criteria has emerged from an inappropriate citation of the source material. By the same token, Harzing (2002) examined the citation network of 60 references on expatriate failure rates (EFRs)—the phenomenon of an expatriate returning home before their contractual period of employment expires. He finds that one reference was cited 22 times, with only six correctly representing the EFRs reported in the original work. Similarly, Todd et al. (2010) report that of 198 randomly selected citations to articles in two recent issues of 33 marine biology journals, 1 in 4 represented an inappropriate citation—ambiguous, no support, or empty (citation to a secondary source).

We find that citations to problematic research are only modestly impacted by the publication of retraction notifications. We also find that the rate at which researchers cite problematic articles without acknowledging problems after the problem has been publicized remains high.

References

Anderson, M. S., Shaw, M. A., Steneck, N. H., Konkle, E. and Kamata, T. (2013) Research integrity and misconduct in the academic profession. In *Higher education: handbook of theory and research*, 217–261. Springer.

Bar-Ilan, J. and Halevi, G. (2017) Post retraction citations in context: A case study. *Scientometrics*, **113**, 547–565.

— (2018) Temporal characteristics of retracted articles. *Scientometrics*, **116**, 1771–1783.

Basu, P. (2006) Where are they now? *Nature medicine*, **12**, 492–493.

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

Broockman, D., Kalla, J. and Aronow, P. (2015) Irregularities in lacour (2014). *Work. pap., Stanford Univ. http://stanford. edu/ dbroock/broockman_kalla_aronow_lg_irregularities. pdf.*

Budd, J. M., Coble, Z. C. and Anderson, K. M. (2011) Retracted publications in biomedicine: Cause for concern. In *Association of College and Research Libraries Conference*, 390–5.

Budd, J. M., Sievert, M. and Schultz, T. R. (1998) Phenomena of retraction: reasons for retraction and citations to the publications. *JAMA*, **280**, 296–297.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. and Munafò, M. R. (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, **14**, 365–376.

Chang, H., Do, B.-R., Che, J.-H., Kang, B.-C., Kim, J.-H., Kwon, E., Kim, J.-Y. and Min, K.-H. (2013) Safety of adipose-derived stem cells and collagenase in fat tissue preparation. *Aesthetic plastic surgery*, **37**, 802–808.

Clarivate (2021a) Web of science. URL: https://www.webofknowledge.com.

— (2021b) Web of science: Summary of coverage. URL: https://clarivate.libguides.com/webofscienceplatform/coverage.

Cokol, M., Iossifov, I., Rodriguez-Esteban, R. and Rzhetsky, A. (2007) How many scientific papers should be retracted? *EMBO reports*, **8**, 422–423.

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*, **100**, 184.

Deer, B. (2011) How the case against the mmr vaccine was fixed. *BMJ*, **342**, c5347.

Erikson, M. G. and Erlandson, P. (2014) A taxonomy of motives to cite. *Social Studies of Science*, **44**, 625–637.

Fanelli, D. (2009) How many scientists fabricate and falsify research? a systematic review and meta-analysis of survey data. *PloS one*, **4**, e5738.

Firinci, F., Karaman, M., Baran, Y., Bagriyanik, A., Ayyildiz, Z. A., Kiray, M., Kozanoglu, I., Yilmaz, O., Uzuner, N. and Karaman, O. (2011) Mesenchymal stem cells ameliorate the histopathological changes in a murine model of chronic asthma. *International immunopharmacology*, **11**, 1120–1126.

Franco, A., Malhotra, N. and Simonovits, G. (2014) Publication bias in the social sciences: Unlocking the file drawer. *Science*, **345**, 1502–1505.

De la Fuente, R., Bernad, A., Garcia-Castro, J., Martin, M. C. and Cigudosa, J. C. (2010) Retraction: Spontaneous human adult stem cell transformation. *Cancer Res*, **70**, 6682.

Godlee, F., Smith, J. and Marcovitch, H. (2011) Wakefield's article linking mmr vaccine and autism was fraudulent. *BMJ*, **342**, c7452.

Grieneisen, M. L. and Zhang, M. (2012) A comprehensive survey of retracted articles from the scholarly literature. *PloS one*, **7**, e44118.

Hamilton, D. G. (2019) Continued citation of retracted radiation oncology literature—do we have a problem? *International Journal of Radiation Oncology* Biology* Physics*, **103**, 1036–1042.

Harzing, A.-W. (2002) Are our referencing errors undermining our scholarship and credibility? the case of expatriate failure rates. *Journal of Organizational Behavior*, **23**, 127–148.

Heibi, I. and Peroni, S. (2021) A qualitative and quantitative analysis of open citations to retracted articles: the wakefield 1998 et al.'s case. *Scientometrics*, **126**, 8433–8470.

Ioannidis, J. P. (2005) Why most published research findings are false. *PLoS Med*, **2**, e124.

Kosaka, H., Ichikawa, T., Kurozumi, K., Kambara, H., Inoue, S., Maruo, T., Nakamura, K., Hamada, H. and Date, I. (2012) Therapeutic effect of suicide gene-transferred mesenchymal stem cells in a rat model of glioma. *Cancer gene therapy*, **19**, 572.

Lance, C. E., Butts, M. M. and Michels, L. C. (2006) The sources of four commonly reported cutoff criteria: What did they really say? *Organizational research methods*, **9**, 202–220.

Lee, T. C., Senecal, J., Hsu, J. M. and McDonald, E. G. (2021) Ongoing citations of a retracted study involving cardiovascular disease, drug therapy, and mortality in covid-19. *JAMA Internal Medicine*, **181**, 1535–1537.

Levelt, W. J., Drenth, P. and Noort, E. (2012) Flawed science: The fraudulent research practices of social psychologist diederik stapel.

Lin, J., Zhang, L. and Yang, H. (2013) Perioperative administration of selective cyclooxygenase-2 inhibitors for postoperative pain management in patients after total knee arthroplasty. *The Journal of arthroplasty*, **28**, 207–213.

Luwel, M., van Eck, N. J. et al. (2018) The schön case: Analyzing in-text citations to papers before and after retraction. In *23rd International Conference on Science and Technology Indicators (STI 2018), September 12-14, 2018, Leiden, The Netherlands*. Centre for Science and Technology Studies (CWTS).

Marofi, F., Alexandrovna, K. I., Margiana, R., Bahramali, M., Suksatan, W., Abdelbasset, W. K., Chupradit, S., Nasimi, M. and Maashi, M. S. (2021) Mscs and their exosomes: a rapidly evolving approach in the context of cutaneous wounds therapy. *Stem Cell Research & Therapy*, **12**, 1–20.

Martinson, B. C., Anderson, M. S. and De Vries, R. (2005) Scientists behaving badly. *Nature*, **435**, 737–738.

McNutt, M. (2015) Editorial retraction. *Science*, aaa6638.

Mena, J. D., Ndoye, M., Cohen, A. J., Kamal, P. and Breyer, B. N. (2019) The landscape of urological retractions: the prevalence of reported research misconduct. *BJU international*, **124**, 174–179.

Panahi, S. and Soleimanpour, S. (2021) The landscape of the characteristics, citations, scientific, technological, and altmetrics impacts of retracted papers in hematology. *Accountability in Research*, 1–16.

Paul, H. Y., Haughom, B. D. and Hansen, E. N. (2015) Comment on" perioperative administration of selective cyclooxygenase-2 inhibitors for postoperative pain management in patients after total knee arthroplasty". *The Journal of arthroplasty*, **30**, 718.

Pfeifer, M. P. and Snodgrass, G. L. (1990) The continued use of retracted, invalid scientific literature. *JAMA*, **263**, 1420–1423.

Rekdal, O. B. (2014) Academic urban legends. *Social Studies of Science*, **44**, 638–654.

RetractionWatch (2021) Retraction watch database. URL: http://retractiondatabase.org/RetractionSearch.aspx?

Rubbo, P., Pilatti, L. A. and Picinin, C. T. (2019) Citation of retracted articles in engineering: A study of the web of science database. *Ethics & Behavior*, **29**, 661–679.

Rubio, D., Garcia-Castro, J., Martín, M. C., de la Fuente, R., Cigudosa, J. C., Lloyd, A. C. and Bernad, A. (2005) Spontaneous human adult stem cell transformation. *Cancer research*, **65**, 3035–3039.

Schneider, J., Yi, D., Hill, A. M. and Whitehorn, A. S. (2020) Continued post-retraction citation of a fraudulent clinical trial report, eleven years after it was retracted for falsifying data. *Under submission to Scientometrics Special Issue on "Bibliometrics and Information Retrieval.*

Service, R. F. (2003) Scientific misconduct. more of bell labs physicist's papers retracted. *Science (New York, NY)*, **299**, 31.

da Silva, J. A. T. and Bornemann-Cimenti, H. (2017) Why do some retracted papers continue to be cited? *Scientometrics*, **110**, 365–370.

Singh, H. P., Mahendra, A., Yadav, B., Singh, H., Arora, N. and Arora, M. (2014) A comprehensive analysis of articles retracted between 2004 and 2013 from biomedical literature–a call for reforms. *Journal of traditional and complementary medicine*, **4**, 136–139.

Steen, R. G. (2011) Retractions in the scientific literature: is the incidence of research fraud increasing? *Journal of medical ethics*, **37**, 249–253.

Steen, R. G., Casadevall, A. and Fang, F. C. (2013) Why has the number of scientific retractions increased? *PLoS One*, **8**, e68397.

Stewart, W. W. and Feder, N. (1987) The integrity of the scientific literature. *Nature*, **325**, 207–14.

Titus, S. L., Wells, J. A. and Rhoades, L. J. (2008) Repairing research integrity. *Nature*, **453**, 980–982.

Todd, P. A., Guest, J. R., Lu, J. and Chou, L. M. (2010) One in four citations in marine biology papers is inappropriate. *Marine Ecology Progress Series*, **408**, 299–303.

Torsvik, A., Røsland, G. V., Svendsen, A., Molven, A., Immervoll, H., McCormack, E., Lønning, P. E., Primon, M., Sobala, E., Tonn, J.-C. et al. (2010) Spontaneous malignant transformation of human mesenchymal stem cells reflects cross-contamination: putting the research field on track–letter. *Cancer research*, **70**, 6393–6396.

Van Der Vet, P. E. and Nijveen, H. (2016) Propagation of errors in citation networks: a study involving the entire citation network of a widely cited paper published in, and later retracted from, the journal nature. *Research integrity and peer review*, **1**, 1–10.

Wadhwa, R. R., Rasendran, C., Popovic, Z. B., Nissen, S. E. and Desai, M. Y. (2021) Temporal trends, characteristics, and citations of retracted articles in cardiovascular medicine. *JAMA Network Open*, **4**, e2118263–e2118263.

Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D., Malik, M., Berelowitz, M., Dhillon, A. P., Thomson, M. A., Harvey, P. et al. (1998) Retracted: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, **351**, 637–641.

Wallis, C. (1983) Fraud in a harvard lab. *Time*, **121**, 49–49.

Supporting Information

SI 1 Classifying Citations as Post-Retraction or Not

A few retraction notices, retracted articles, and articles citing the retracted articles have earlier online (or conference) publication dates than the print publication dates recorded by WoS. As a result, post-retraction citations can be classified as otherwise. Or vice versa. To determine the impact of this issue and issues like these on our estimate of the lower bound of the proportion of citations that are made a year or more after the publication of the retraction notice, we manually recorded the online publication dates for a random sample of 300 citations to retracted articles, the associated retraction notices, and retracted articles. We could not retrieve 20 articles citing a retracted article. Of the remaining 280 records, switching to online publication dates suggests that three articles were misclassified as post-retraction (2.2%) and four were misclassified as not post-retraction (3.2%). Taking these error rates at their face value, we re-calibrated our results. The recalibration results in an increase in the number of post-retraction citations, from 22,932 to 23,289. Or, the lower bound of the proportion of citations that happen the year after the retraction notice is published goes from 31.2% to 31.7%.

SI 2 Coding Citations as Acknolwedging Problems Or Not

To code the citations, we downloaded citing articles and their associated retracted article. A research assistant then edited the citing article pdf to highlight where the retracted article was discussed in the citing article. The judgment of whether the article noted any concerns was made based on a review of the original retracted article pdf and the highlighted text.

If an article did not note any concerns with the cited article, it was coded as *not acknoledging problems*. Simply disagreeing with the conclusions of an article without noting any concern still meant that the article was being cited in a way that suggests that its findings are trustworthy and were also coded as *not acknowledging problems*. We code articles that note any concern with the citing article, even those unrelated to the cause of retraction, as *acknowledging problems*.

In the retracted article data, we could not locate 32 articles, leaving us with 343 articles. There were no false positives. Of the 87 articles citing retracted articles before or in the same year the retraction notice was published, 85 (97.7%) did not acknowledge problems. And of the 256 articles citing retracted articles the year after the retraction notice was published, 234 (91.4%) did not acknowledge problems.

We evaluated the reliability of the coding by having an independent rater code 50 randomly selected citing articles. The two sets of independent codes were found to agree in all 50 instances.