

Propagation of Error: Approving Citations to Problematic Research

Ken Cor,¹ Gaurav Sood^{2*}

¹Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton

²Principal Data Scientist, Microsoft, Washington

*Corresponding author. E-mail: gasood@microsoft.com.

Reports of serious errors in published research are increasingly common. Once the issues have been made public, we expect *approving* citations to the problematic articles—citations noting no concerns with the cited article—to stop. Using a novel database of over 3,000 retracted articles and nearly 74,000 citations to the retracted articles as well as data from a prominent article that highlights a statistical error in a set of articles published in prominent journals, we estimate citation rates and rates of approving citations pre- and post-notification. We find that at least 31.2% of the citations to retracted articles happen a year after the article has been retracted. And that 91.4% of these post-retraction citations are *approving*. We also find that problematic research continues to be approvingly cited long after the problems have been publicized. Our results have implications for the design of scholarship discovery systems and scientific practice more generally.

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 convey that the results can be trusted.

When researchers *approvingly* cite—cite without noting any concerns—articles with serious errors, problems ensue. First, when such articles (e.g. (1) as noted by (2)) are approvingly cited as evidence, e.g., (3), it cues that the evidence for the claim is good. Such citations thus unduly increase the reader’s confidence in the result or argument. When the claim being buffeted by the citation is wrong, such citations also misinform. And a misinformed reader may propagate the error further by sharing the point with colleagues and students or by publishing research with the erroneous claim.

Second, approving citations to erroneous studies gives 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.

Third, when erroneous research (e.g. (1)) is approvingly cited to contextualize results, e.g., (4), readers get a misleading impression of the plausibility of the numbers reported in the study.

Fourth, sometimes data from retracted articles are used in meta-analyses. For instance, as noted by (5), (6) use data from two retracted articles in their meta-analysis, which since then has been approvingly cited multiple times (e.g. (7)).

In sum, *approving* citations to problematic research propagate the error. But we expect publication of an article highlighting the problems, such as a retraction notice, to stem the propagation. We expect approving citations to problematic research to stop once the problems have been made public. Some research, however, suggests otherwise. For instance, research

by John Darsee continued to be approvingly cited after a highly publicized retraction of his work (8). Similarly, a study using a database of 235 retracted biomedical articles found that nearly 94% of the citations after retraction treated research as valid (9). Yet another study, exploiting a dataset of 82 retracted articles, came to similar conclusions (10).

All these studies, however, suffer from three weaknesses. First, the studies use small samples, often spanning a single discipline. Use of small, selective samples means that we still do not know how widespread the problem is. Second, the studies solely focus on citations after retraction. This means that we do not know how common approving citations are before the problems are publicized, and if the rate changes after publicity. And third, the studies only focus on retracted articles. Retraction is generally a result of serious scientific malpractice. By focusing on retractions alone, the studies fail to illuminate the much more common problem of approving citations to studies with major errors with potentially serious implications for the key results.

We address all three issues. We study approving citations to articles that make potentially serious errors that don't lead to a retraction by leveraging data from an article published in *Nature* that highlights a potentially serious statistical error in articles published in prominent journals. To more extensively study approving citations to retracted articles, we assemble a large original dataset of retracted articles—over 3,000 retracted articles and nearly 74,000 citations to the retracted articles.

Our first dataset comes from articles that mistake the difference between a statistically significant and statistically insignificant result as evidence that the difference is statistically significant (11). (For why this is problematic, see (12).) Nieuwenhuis et al. (11) analyze 170 articles published in *Nature*, *Science*, *Neuron*, and *Journal of Neuroscience* between 2009 and 2010. They find that roughly half of the 170 articles conducting such an analysis make the mistake. We augment their data by using Web of Science (WoS) (13) to download citation records of all

170 articles.

Using these data, we study the frequency of citations and the frequency of approving citations before and after the error is made public. We expect publication of an article noting the statistical error to increase awareness about the error, and to reduce the frequency of citations, especially approving citations, to articles with the error.

Between 2009 and 2011, the articles making the mistake were cited 2,267 times. Between 2012 and 2015, the articles were cited an additional 6,604 times. Fig. 1 plots the number of citations received by articles making the mistake, the average number of citations received per year, and the smoothed `loess growth` curves. As the figure shows, the average number of citations increases roughly smoothly. To account for the skew in citations, in fig. S1, we switch means with medians. Switching to medians yields a similar pattern except for the expected intercept shift.

Not all errors are equally consequential for the results. Nieuwenhuis et al. flag articles where the error has potentially serious consequences for the results. Citations to these articles show a similar pattern of increasing citation rates over time (see Fig. S1). Comparing citations to articles making the mistake to citations to articles published in the same venue but not making the mistake using a difference-in-difference regression suggests articles making the mistake are cited *more often* (see S3).

But citations to erroneous research do not need to decline after the error in the research is publicized. We only expect approving citations to decline. To estimate how many citations after the publication of Nieuwenhuis et al. are approving, we coded whether the citation was approving or not in 100 randomly chosen articles citing articles with the mistake (see SM 1.4 for details about the coding.) Of the 100, only one article noted concerns with the cited article, citing (11) for support. In all, there is strong evidence that approving citations are extremely common after the error is publicized.

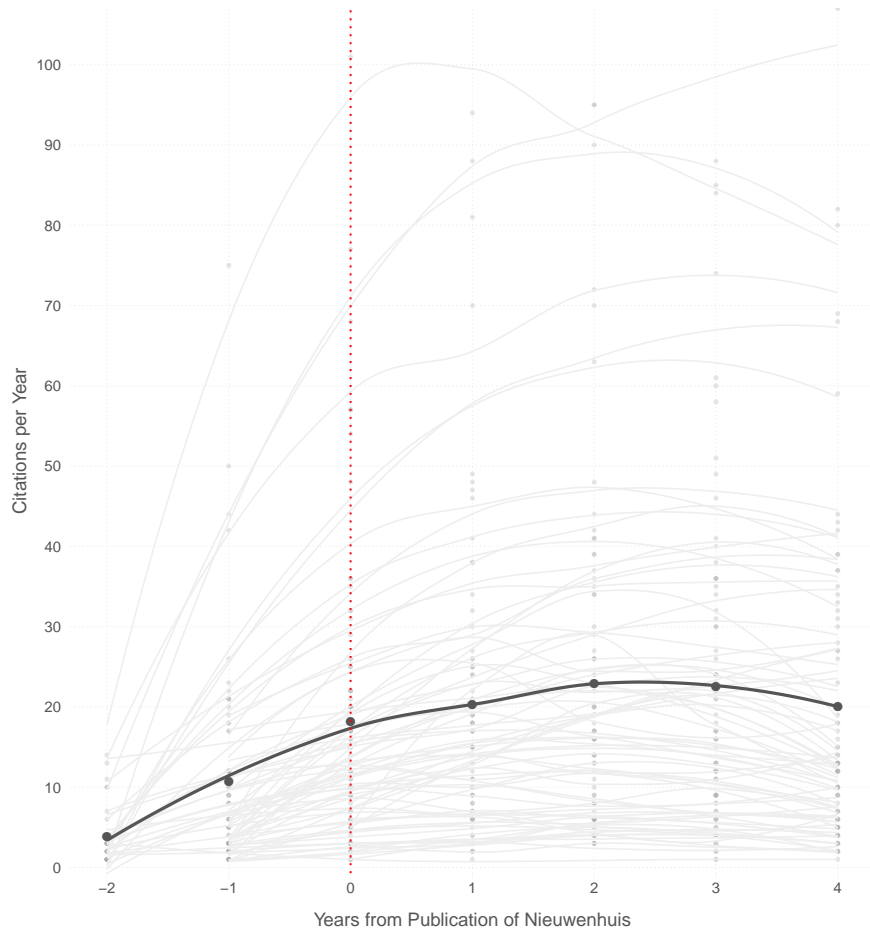


Figure 1: Number of Citations to Articles Containing the Error Per Year.

We expected approving citations to articles making a statistical error to decline, but our expectations were tempered by the fact that publications about statistical errors do not gain the same publicity as retractions. To study how citations to retracted articles fare before, we leveraged a large novel dataset of retracted articles and citations to them.

We used WoS to assemble the data. WoS indexes articles from over 9,500 natural science journals and 3,500 social science journals (14). (See Supplementary Materials (SM) 1.1 for details about the WoS database.) To assemble the data, we searched WoS for retraction notices, used information in the retraction notice records to download retracted articles, and then down-

loaded citations to the retracted articles using WoS feature that allows you to access citations associated with an article. (For details about the method and robustness checks around data, see SM 1.2.)

Our final dataset has 3,029 retracted articles and 73,564 citations to the retracted articles. 65.2% of the retracted articles are from Life Sciences and Biomedicine and 13% are from Physical Sciences. (For a breakdown by field, see Table S1.) Data also show that the number of retractions has increased sharply over the last thirty years (see Fig. S3). Between 1989 and 1999, the number of retraction notices being published per year never crossed 20. Since then, there has been a sharp rise in the number of retraction notices per year. Between 2001, when 15 retraction notices were published, and 2015, the last year for which we have complete data, there was a near 30 fold increase when a total of 439 retraction notices were published. The pattern that we find is consistent with results from other studies looking at retraction rates (15).

The rapid rise in the number of retractions is likely a combination of increasing production and improvements in detection. But the bottom line is that there is an ever faster growing number of articles that the scientific community thinks 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. Of the 100, 39% 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 retractions, with fraud alone contributing to 24% of 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 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 (16) (see

also (17)). All in all, articles are mostly retracted because the research cannot be trusted.

These flawed articles often accrue a fair number of citations before being retracted. In our data, the articles had been cited 39,792 times before being retracted. 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 Fig. 2) with 28.1% of the articles taking 4 or more years to be retracted. These numbers are similar to those obtained elsewhere. A study on time to retraction in the PubMed corpus found that the average time to retraction was approximately three years (15).

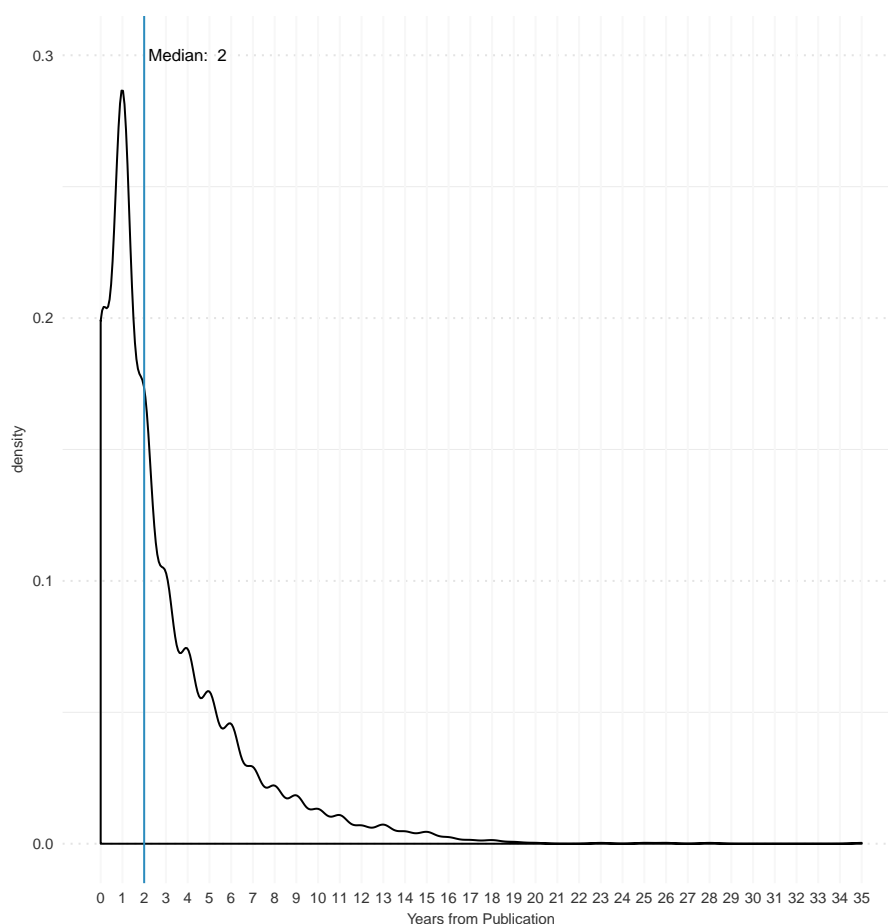


Figure 2: Time to Retraction

On the hunch that greater readership of more prominent journals would mean that problematic articles are flagged more quickly, we estimated the relationship between journal impact factor and average time to retraction. As Fig. S4 shows, the relationship is flat—flawed articles in low ranked journals are retracted as quickly as flawed articles in higher ranked journals.

Given that a majority of retracted articles are retracted because of serious error or fraud, we expect retracted articles to *never* be approvingly cited 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. Thus, on average, the retracted articles received an additional 7.57 citations. Given the skew in retraction notices, with a 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 happen after an article has been retracted.

To explore the frequency of citation before and after retracted, we plotted line graphs of total citations per article per year against year from the publication of retraction notice (see Fig. 3). And we overlaid the lines with the median number of citations per article per year. We limit ourselves to 10 years before and after the publication of retraction notice as the data are very sparse beyond that. The number of citations decline when the retraction notice is published (the median goes from 3 to 2 between the year retraction notice is published after next year) but is followed by a plateauing. Regressions that control for a time trend show the same (see Table S4). Citations continue to accrue at a steady low rate long after the article has been retracted.

To estimate how many of the citations are *approving*, we 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. 97.7% of the articles cited retracted research approvingly before or in the same year as the retraction was published. The rate drops to 91.4% for articles citing retracted articles a year or more after retraction. (Fig. 4 plots pro-

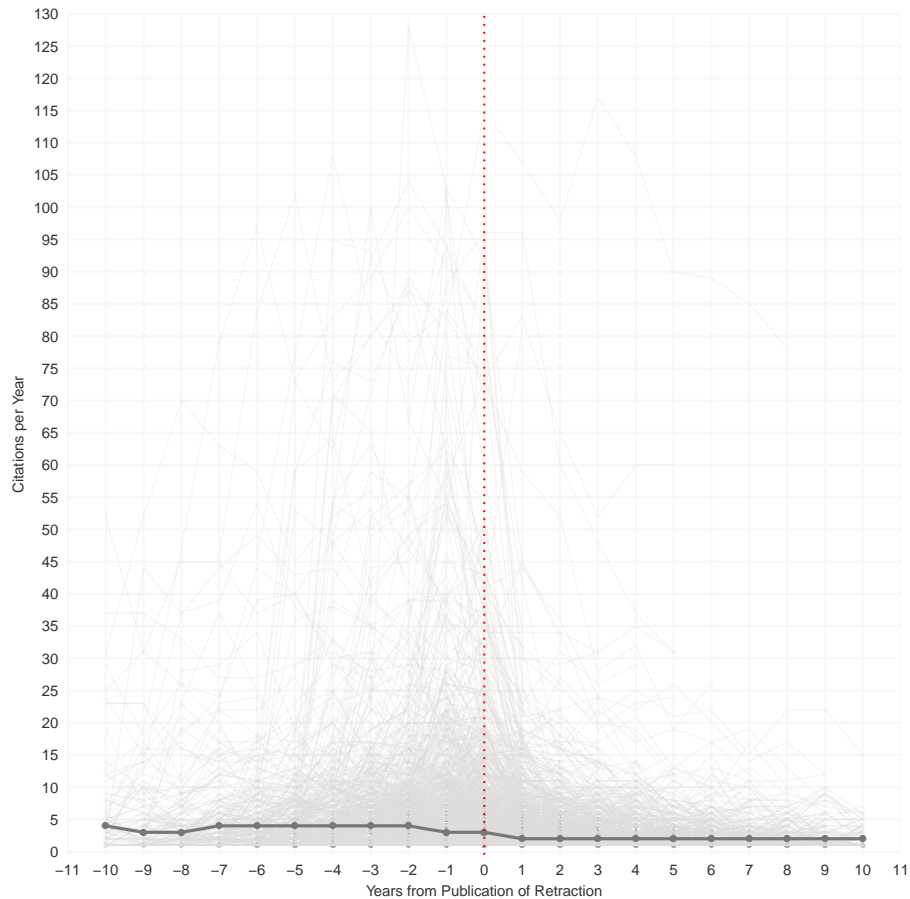


Figure 3: Number of Citations to Retracted Articles per Year

portion of citations approving by year.) In all, data suggest that problematic research continues to be approvingly cited at a high rate long after problems have been publicized.

But why do researchers approvingly cite problematic research even after the problems have been publicized? It is likely because they are unaware of the problems. That is not the same as saying that they couldn't have known, but that there are still barriers to finding out. For instance, Google Scholar does not flag if an article has been retracted. Similarly, bibliographic software does not alert researchers that problems have been discovered about a paper in their bibliographic database (18).

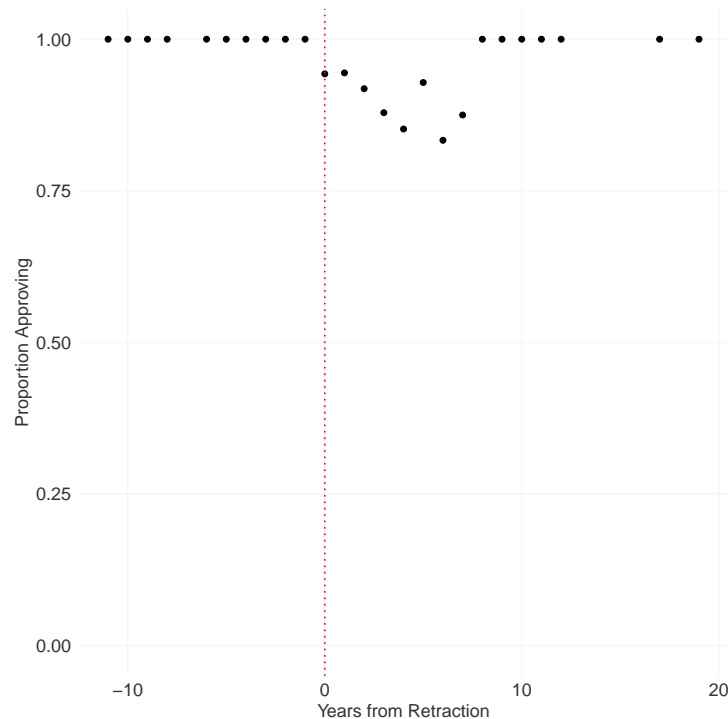


Figure 4: Proportion of Citations That Are Approving Per Year

The cost of finding out is not zero but neither is it too high. Many academic publishers flag retracted articles with the prefix “retracted.” We think that the reason scientists do not mount the low barriers—spend time carefully vetting each article—is because serious pressures to publish leave them with little time and because they trust published research.

There are two reasons why we think scientists trust published research. The first is that the rate of retractions isn't extremely low. For instance, of the nearly 9.4 million articles published between 1950 and 2004 and available on PubMed, only 596 have been retracted (19). In all likelihood, however, the true rate of serious errors in manuscripts is manifolds that rate. The same study estimates the rate at which articles ought to be retracted to be anywhere between 16.7 times to 167.8 times the actual rate.

The other likely reason why scientists trust published research is that they believe that sci-

entific misconduct is limited to a few people. And that optimism is likely driven by the fact that a few cases of fraud get a bulk of the attention, with reporting often focusing on personalities rather than processes. Cases like Diederik Stapel, who fabricated data behind at least 30 papers (20), John Darsee, who faked data behind nearly 100 publications (21, 22, 23), and Jan Hendrik Schön, who during a period in 2001 published a research paper every 8 days based on fabricated data (24, 22), Andrew Wakefield, who published a paper linking MMR vaccine to autism using fabricated data (25, 26, 27), and Michael Lacour, who published a paper in *Science* based on fabricated data (28, 29) get a bulk of the attention (see also (30, 22)).

But misconduct is not limited to a few bad actors. A large anonymous survey of early- and mid-career scientists found that nearly 2% of the scientists reported fabricating, falsifying, or plagiarizing in the last *three* years (31) (see also (32)). Another study found that nearly 34% of the respondents admitted to engaging in questionable research practices (33).

In all, researchers continue to approvingly cite problematic research because the search costs are relatively high. To ameliorate the issue, we need to build tools that provide information during the research discovery and production processes. For instance, a browser plug-in that highlights problematic articles listed on a web page is likely to be useful. Providing such a tool to editors at academic publishers can help ameliorate the problem.

References

1. D. Rubio, *et al.*, *Cancer research* **65**, 3035 (2005).
2. A. Torsvik, *et al.*, *Cancer research* **70**, 6393 (2010).
3. H. Chang, *et al.*, *Aesthetic plastic surgery* **37**, 802 (2013).
4. H. Kosaka, *et al.*, *Cancer gene therapy* **19**, 572 (2012).
5. H. Y. Paul, B. D. Haugom, E. N. Hansen, *The Journal of arthroplasty* **30**, 718 (2015).
6. J. Lin, L. Zhang, H. Yang, *The Journal of arthroplasty* **28**, 207 (2013).
7. M. W. Russo, N. L. Parks, W. G. Hamilton, *Orthopedic Clinics* **48**, 401 (2017).
8. C. A. Kochan, J. M. Budd, *Journal of the American Society for Information Science* **43**, 488 (1992).
9. J. M. Budd, M. Sievert, T. R. Schultz, *JAMA* **280**, 296 (1998).
10. M. P. Pfeifer, G. L. Snodgrass, *JAMA* **263**, 1420 (1990).
11. S. Nieuwenhuis, B. U. Forstmann, E.-J. Wagenmakers, *Nature neuroscience* **14**, 1105 (2011).
12. A. Gelman, H. Stern, *The American Statistician* **60**, 328 (2006).
13. T. Reuters (2012).
14. J. Yong-Hak, *Thomson Reuters* (2013).
15. R. G. Steen, A. Casadevall, F. C. Fang, *PLoS One* **8**, e68397 (2013).

16. J. M. Budd, Z. C. Coble, K. M. Anderson, *Association of College and Research Libraries Conference* (2011), pp. 390–5.
17. R. G. Steen, *Journal of medical ethics* pp. jme–2010 (2010).
18. P. M. Davis, *Journal of the Medical Library Association* **100**, 184 (2012).
19. M. Cokol, I. Iossifov, R. Rodriguez-Esteban, A. Rzhetsky, *EMBO reports* **8**, 422 (2007).
20. W. J. Levelt, P. Drenth, E. Noort (2012).
21. W. W. Stewart, N. Feder, *Nature* **325**, 207 (1987).
22. M. S. Anderson, M. A. Shaw, N. H. Steneck, E. Konkle, T. Kamata, *Higher education: handbook of theory and research* (Springer, 2013), pp. 217–261.
23. C. Wallis, *Time* **121**, 49 (1983).
24. R. F. Service, *Science (New York, NY)* **299**, 31 (2003).
25. A. J. Wakefield, *et al.*, *The Lancet* **351**, 637 (1998).
26. B. Deer, *BMJ* **342**, c5347 (2011).
27. F. Godlee, J. Smith, H. Marcovitch, *BMJ* **342**, c7452 (2011).
28. D. Broockman, J. Kalla, P. Aronow, *Work. pap., Stanford Univ.* http://stanford.edu/dbroock/broockman_kalla_aronow_lg_irregularities.pdf (2015).
29. M. McNutt, *Science* p. aaa6638 (2015).
30. P. Basu, *Nature medicine* **12**, 492 (2006).
31. B. C. Martinson, M. S. Anderson, R. De Vries, *Nature* **435**, 737 (2005).

32. S. L. Titus, J. A. Wells, L. J. Rhoades, *Nature* **453**, 980 (2008).
33. D. Fanelli, *PloS one* **4**, e5738 (2009).

Acknowledgments

Data and replication code are available on GitHub at http://github.com/recite/propagation_of_error.

Both authors contributed equally to all aspects of the research. No funding was required for this article. The authors declare no conflicts of interest. We are grateful to Danielle Portnoy and Xiaoran Huang for assisting us with research, and to Andrew Gelman, Kabir Khanna, and Daniel Stone for offering valuable comments.

Supplementary Materials

This PDF file includes:

- Materials and Methods
- Supplementary Text
- Fig. S1-S4
- Tables S1-S4

1 Materials and Methods

1.1 Web of Science Indices

WoS indexes articles from over 12,000 international journals and 148,000 conferences (14). WoS contains key citation indices including the *Science Citation Index Expanded* (over 9,500 journals; 1900–present), *Social Sciences Citation Index* (over 3,500 journals; 1900–present), *Arts & Humanities Citation Index* (over 1,700 journals; 1975–present), *Conference Proceedings Citation Index* (over 170,000 conferences; 1990–present), *Book Citation Index* (over 30,000 titles; 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*, see https://github.com/recite/propagation_of_error/data/11_wos/what_is_in_wos/.

1.2 Steps for Creating the Retractions Dataset

To build a 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. 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.

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 us to 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 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. (It is standard practice for titles of original articles to be revised to indicate that the article has been retracted.) 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. And that served as our final sample.

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.

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 automatically download citation records for each of the retracted articles. In total, we found 73,564 citations.

1.3 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%.

1.4 Retracted Articles by Field

To understand the kinds of fields where retractions are more common, we used an augmented Web of Science research field categorization scheme to classify the articles. There is one caveat. Sometimes papers cover more than one topic. We just choose the first topic in these cases taking it to be the primary topic.

As Table S1 shows, 65% of the retracted articles were published in the Life Sciences and Biomedicine field. A distant second at 13% is Physical Sciences, followed by Technology at

10.7%. Social Sciences are 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.

Coding Citations as Approving or Not

To estimate what proportion of the citations are approving, we coded a random sample. For Nieuwenhuis, we subset on citations to articles making the error after the publication of Nieuwenhuis, and then randomly chose a 100. For the retracted article dataset, we picked 375 citations using stratified random sampling so that we had 100 pre-retraction citations and 275 post-retraction citations.

To code the citations, we downloaded the citing 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 does not note any concerns with the cited article, it implicitly *approves* of the findings. Simply disagreeing with the conclusions of an article without noting any concern or citing the article in a he-said-she-said way still mean that the article is being cited in a way that suggests that its findings are trustworthy. So we code such instances as *approving*. We code articles that note any concern with the citing article, even those unrelated to the cause of retraction, as *disapproving*.

We evaluated the reliability of the coding by having an independent rater code 50 randomly selected retracted articles. The two raters gave the same labels to all 50 articles. We take this as evidence that the coding was reliable.

In the Nieuwenhuis data, we could not locate one of the 100 articles, leaving us with 99 articles. Of the 99 articles, 2 were false positives—the articles did not cite erroneous research,

but instead cited a paper with authors and title similar to published erroneous research. Of the 97 remaining articles, only one article noted concerns while citing an article making a mistake, citing Nieuwenhuis et al. (11) for support.

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%) were approving. And of the 256 articles citing retracted articles the year after the retraction notice was published, 234 (91.4%) were approving.

2 Supplementary Text

2.1 Impact of Publication of Nieuwenhuis et al.

In figure 1, we had plotted the total number of citations received per year by each of the papers making the mistake, the average number of citations received per year by articles making the mistake and smoothed `loess` growth curves in the manuscript. But the plot suggests that citation rates are skewed. To account for the skew, we switched means with medians (see Fig. S1). Doing so yields a similar pattern except for the expected intercept shift.

Nieuwenhuis et al. (11) flag articles where the error has potentially serious consequences for the results. So we subset on such articles and plot how the median number of citations varies across years. As Fig. S2 shows, the median number of citations steadily and modestly increase over time.

To more formally explore the change in citation rate as a consequence of the publication of (11), we started with an ‘event study design’ focusing on the citation rates to articles with the error. We regressed citations per year on a dummy for the year (11) was published, a linear time trend, and fixed effect for the article. We also cluster by articles to account for multiple observations per article. In effect, we are getting an average of within article changes

after regressing out a linear time trend. Results show, if anything, a modest uptick in citations after (11) is published—a year after publication of (11), articles containing the error get about four more citations per year compared to what they were getting before it (see Table S2).

Our most complete analysis for the (11) data is a Difference-in-Differences (DID) analysis. DID gives us a better way to control for over time trends, though as you will see, the results echo the results from the simpler analysis. We estimated whether the difference in citation rates of articles making the error and those not making the error changed after the publication of (11). In particular, we regressed citations per year on whether or not the article makes the error, the year(s) after the publication of (11), and interaction between the two. We also include fixed effects for each article to do within article estimation. This allows us to circumvent concerns around skew in citations. And we again clustered the standard errors by article.

Table S3 tabulates the results. Models (1), (3), and (5) define error as any article making the error. And Models (2), (4), and (6) refer to error as articles making “potentially serious errors.” As the table shows, 1 or 2 years after (11), articles making the error were being cited more frequently vis-à-vis articles not making the error (Diff. ~ 3). Three years out, we cannot still reject 0, suggesting that there is no evidence of a decline. For articles making “potentially serious errors”, the story is much the same, except that the 1 and 2-year out estimates are closer to 3.5 additional citations per year than 3. And three years later, we still cannot say that the articles making “potentially serious errors” were being cited any less frequently.

2.2 Number of Retracted Articles by Year in the WoS Dataset

Fig. S3 shows the number of retractions in the WoS dataset as a function of year of publication. As is clear, the number of retracted articles is increasing exponentially.

The rapid rise in the number of retractions is likely a combination of increasing production and improvements in detection. But the bottom line is that there is an ever faster growing

number of articles which the scientific community thinks should not have been published in the first place. As we show above, articles are not retracted for minor violations, they are mostly retracted for major lapses in the scientific process.

2.3 Relationship Between Journal Impact Factor and Time to Retraction

We estimated the relationship between journal impact factor and average time to retraction. Journals with higher impact factor have no shorter time to retraction than journals with lower impact factors (see Fig. S4).

2.4 Rate of Citations Before And After Retraction

Fig. 3 elides over the fact that we do not observe data on all articles after all the plotted years after the publication notice. For instance, if the retraction notice was published in 2014, we only observe one more full year of citations—2015. Thus, to look more formally at the impact of publication of retraction notices a year, 2 years, and 3 years after, we create subsets of data where we have all the articles for which we have at least 1 year, 2 years, and 3 years worth of data after the publication of the retraction notice. To analyze these subsets, we use a pretty simple model. We regress the number of citations per year per article on years from retraction notice, a dummy for the cliff (1-, 2-, and 3- years after the publication of the retraction notice) and an interaction between the two. We cluster the standard errors by article to account for the fact that we have multiple observations per article. The results of the model can be seen in Table S4.

As Table S4 shows, on average an article is cited about 5–6 times per year. But 1, 2, and three years later, an average article accrues about two less citations per year, a drop that is statistically and substantively significant. There is also a small negative slope in the number of citations per year for models that estimate the effect 1 and two years out. So the number of

citations is slowly decreasing. But note that our priors are post-retraction, articles would not be cited. So we must compare the citation rate to 0. And there we can easily reject the 0—citation rate post publication of the retraction notice is not zero.

3 Tables S1 to S4

Table S1: Retraction Notices By Field

Field	Number of Notices	Percentage of Total
Arts & Humanities	13	0.4
Life Sciences & Biomedicine	1974	65.2
Multidisciplinary	157	5.2
Physical Sciences	393	13.0
Social Sciences	165	5.5
Technology	325	10.7

Table S2: Change in the Number of Citations to Articles Containing the Error Per Year Before and After Publication of Nieuwenhuis

	<i>Dependent variable:</i>	
	Citations Per Year	
	All Articles with Mistakes	Articles with Potentially Serious Errors
	(1)	(2)
Transition Date	3.8** (1.7)	5.0** (2.0)
Time	2.0*** (0.4)	2.1*** (0.5)
Constant	12.4*** (1.9)	11.6*** (2.1)
Observations	487	276
Akaike Inf. Crit.	3,818.0	2,095.8
Bayesian Inf. Crit.	3,838.9	2,113.9

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S3: Difference-in-Difference Analysis of the Impact of Publication of Nieuwenhuis on the Number of Times per Year Articles Containing the Error Are Cited Vis-a-Vis Articles that Didn't Contain the Error

	<i>Dependent variable:</i>					
	1 year out		2 years out		3 years out	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Date	7.2*** (0.9)	7.7*** (0.7)	5.1*** (0.9)	5.5*** (0.7)	3.7*** (1.0)	3.9*** (0.8)
Error or Not	1.5 (2.5)	0.004 (2.8)	2.1 (2.4)	0.6 (2.7)	2.8 (2.4)	1.6 (2.6)
Makes Error*Treatment Date	3.1** (1.2)	3.7*** (1.3)	2.7** (1.2)	3.5*** (1.3)	1.7 (1.3)	2.1 (1.5)
Constant	9.5*** (1.8)	10.2*** (1.5)	11.7*** (1.7)	12.6*** (1.5)	13.1*** (1.7)	14.0*** (1.4)
Observations	957	957	957	957	957	957
Akaike Inf. Crit.	7,328.2	7,327.8	7,408.8	7,407.5	7,474.9	7,475.2
Bayesian Inf. Crit.	7,357.4	7,357.0	7,437.9	7,436.7	7,504.0	7,504.4

Note: *p<0.1; **p<0.05; ***p<0.01

Models (1), (3), and (5) define error as any article making the error.

And Models (2), (4), and (6) refer to error as articles making "potentially serious errors."

Table S4: Impact of Publication of Retraction Notice on the Number of Times Retracted Articles Are Cited per Year

	<i>Dependent variable:</i>		
	Citations Per Year		
	1 Year Later	2 Years Later	3 Years Later
	(1)	(2)	(3)
(1, 2, 3) Years After Notice	−2.4*** (0.2)	−2.1*** (0.2)	−1.9*** (0.3)
Years to Notice	−0.01 (0.03)	−0.2*** (0.03)	−0.3*** (0.03)
(1, 2, 3) Years After Notice*Years to Notice	−0.4*** (0.04)	−0.2*** (0.05)	0.002 (0.1)
Constant	6.0*** (0.2)	5.2*** (0.1)	4.7*** (0.2)
Observations	12,511	11,486	10,428
Akaike Inf. Crit.	83,835.6	76,511.3	69,534.9
Bayesian Inf. Crit.	83,880.2	76,555.4	69,578.4

Note:

*p<0.1; **p<0.05; ***p<0.01

4 Figures S1 to S4

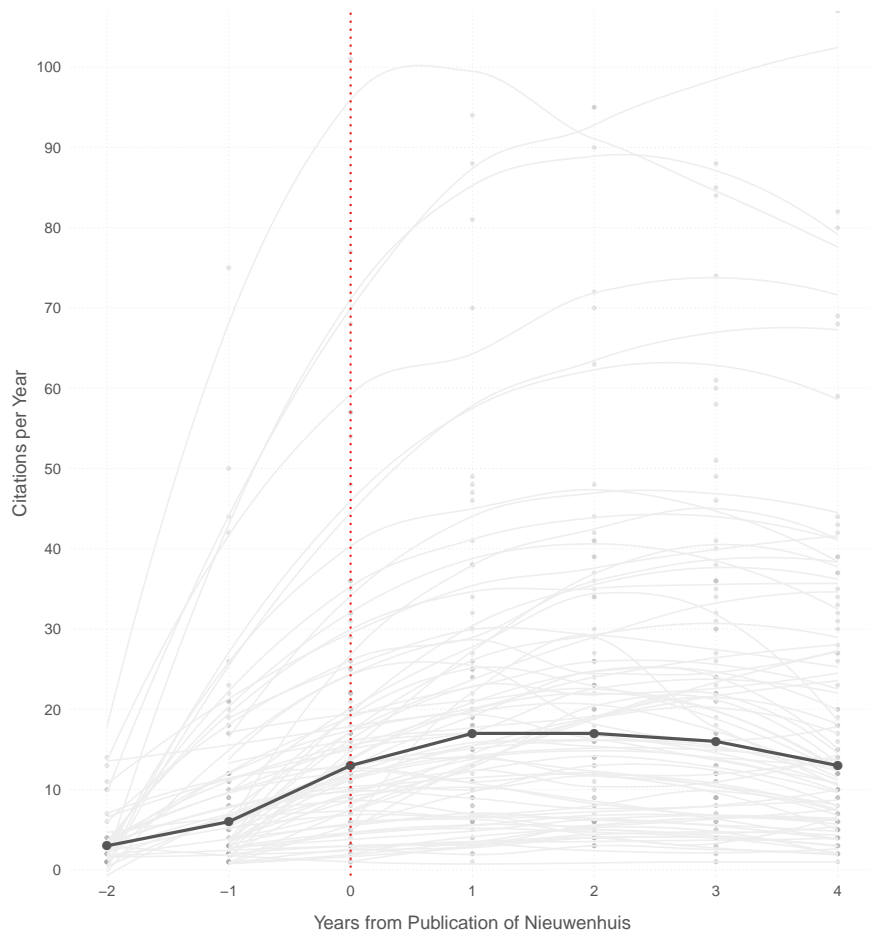


Figure S1: Total number of citations received per year by each of the papers making the mistake, and the median number of citations received per year by the articles.

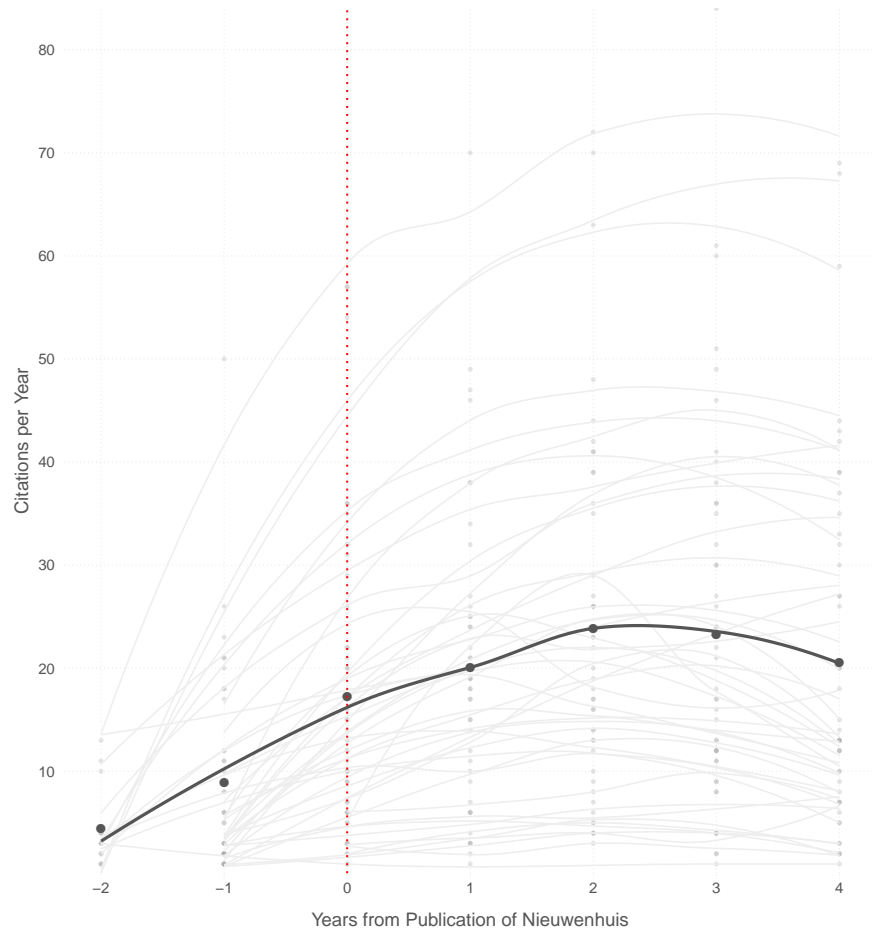


Figure S2: Total number of citations received per year by articles making the mistake with ‘potentially serious’ consequences for the results, and the average number of citations received per year by the articles.

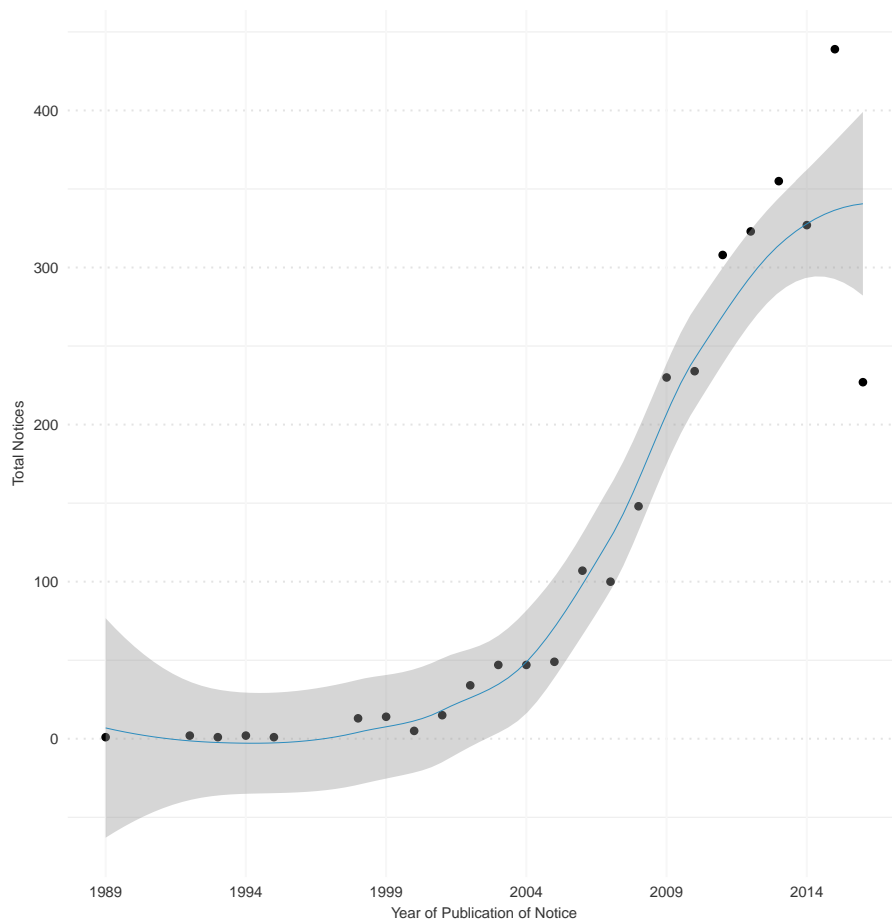


Figure S3: Retraction Notices Per Year

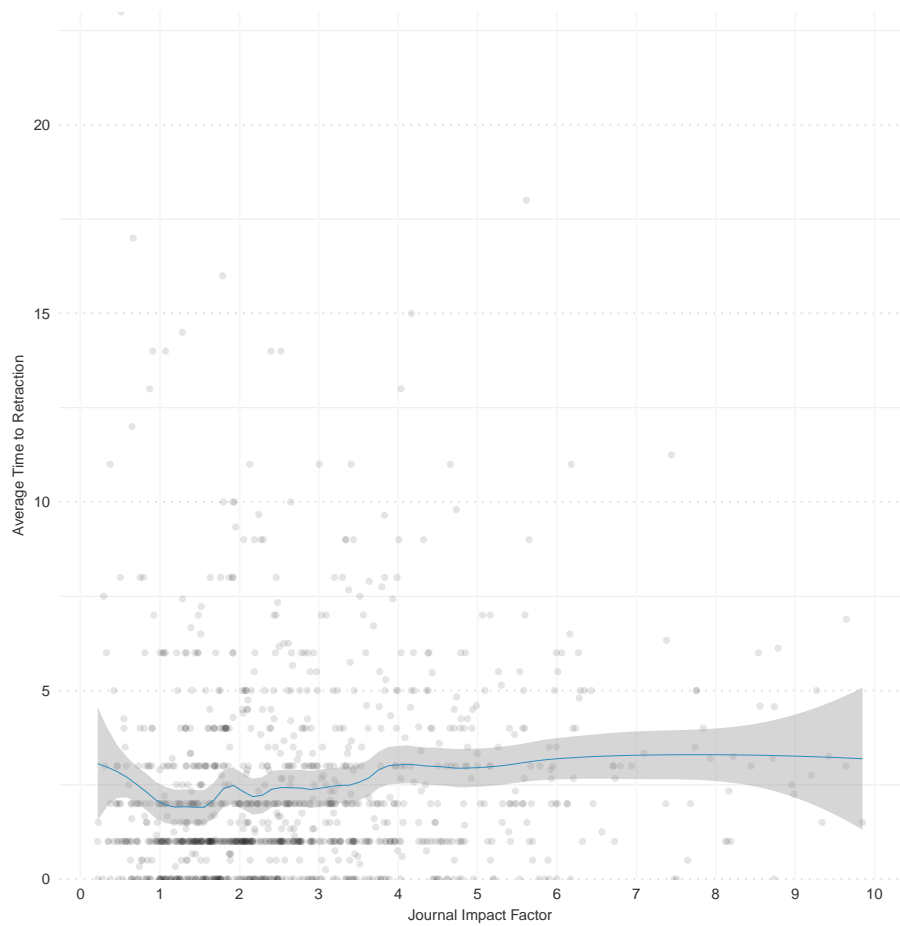


Figure S4: Relationship Between Journal Impact Factor and Time to Retraction