Propagation of Error: Impact of Publicizing Serious Problems on Citations to Problematic Research*

Gaurav Sood† Ken Cor‡

February 28, 2017

NB: Preliminary draft. Please do not cite without permission.

Abstract

Reports of serious errors in academic research are increasingly common. Once the error has been made public, either via a retraction or via publication of research that points to the error, it is often assumed that information about the error has been widely disseminated. And that *approving* citations to the erroneous piece of research will cease. Using a large novel set of retracted articles—over 3,000 retracted articles and over 80,000 citations to retracted articles—and data from a prominent article that highlights a potentially serious concern in a set of articles published in prominent journals, we estimate the change in rate of citations to flawed research due to publicizing the error. We find there is at best a small effect of making errors public on citation counts three years after the error is made public. Our results have implications for design of scholarship discovery systems, and for scientific practice more generally.

*We are grateful to Danielle Portnoy and Xiaoran Huang for their help in research, and to Kabir Khanna and Doug Ahler for offering valuable comments. The data and scripts for replicating the analysis can be found at: https://github.com/soodoku/propagation_of_error

†Gaurav can be reached at gsood07@gmail.com

‡Ken is the Director of Assessment in the Faculty of Pharmacy and Pharmaceutical Sciences at the University of Alberta, Edmonton. Ken can be reached at mcor@ualberta.edu

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" or "our study uses the same method as x"), and to contextualize results ("our results are consistent with results from y"). And unless the researcher notes specific problems with the research being cited, citations cue that the cited research was done in good faith, and that the data, the results, and the inferences are correct.

When researchers "approvingly" cite—cite without mentioning any concerns—erroneous articles, a variety of problems ensue. First, such citations give full credit to research (and researchers) when at best partial credit is deserved. Citation tallies, in turn, are a cue for credibility. And such citations make erroneous research appear credible. Second, approvingly citing erroneous research to support a claim suggests that the evidence for the claim is good. At the very least, they unduly increase readers' confidence in a result or argument. When the claim is wrong, such citations misinform. In the extremum, a reader may become persuaded that the incorrect point is right. And such a reader—generally another academic—may go on to write other articles influenced by the incorrect point, citing the erroneous article for support, or share the point as fact with colleagues and students, propagating the error.

If erroneous research is approvingly cited as evidence, or for contextualizing results, the problems extend to the paper citing the flawed research. The perniciousness of such citations varies by how central the argument buffeted by the flawed article(s) is for the thesis of the paper. In some cases, it is likely enough to upend the entire article. In other cases, consequences may be less dramatic. By the same token, when citation to flawed studies are used to contextualize and support the results—for instance, like x and y, our results show z—plausibility of the numbers from the study are put in doubt.

Given that approving citations to erroneous research can result in serious problems, we study frequency of citations to research with serious errors, and the impact of publicizing serious errors in research on future citations to problematic research. To study the question, we assemble a large original dataset of retracted articles, and separately, leverage data from an article that highlights a potentially serious problem in articles published in prominent journals. Using these two datasets, we first shed light on the scientific retraction process more generally, providing estimates of the time between publication of an article that will eventually be retracted and publication of the retraction notice, frequency of retraction by research area, and reasons why articles are retracted. Next, we estimate the average number of citations problematic articles accumulate before they are retracted. Following that, using an interrupted time series design, we estimate the impact of publication of retraction of an article or publication of an article that highlights serious errors in an article on its citations. Data suggest that even when serious misconduct is made public via retractions, approving citations to problematic research remain common.

Why is Research With Serious Errors Approvingly Cited?

Perhaps the single most important reason why researchers approvingly cite research with serious errors, even after the errors have been publicized, is that is the pressure to publish. Given the pressure to publish, many researchers likely do not spend enough time vetting the research they cite.[1] Pressed for time, scientists often likely default to credulousness when evaluating the research they cite. Though there is also likely some 'motivated vetting,' with articles cited in 'support' likely receiving less scrutiny than those making an 'opposing' argument.

The second most important reason is scientists trust other scientists, especially peer-reviewed work produced by other scientists. One likely reason for that trust is the belief that scientific misconduct is limited to a few bad 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 than processes. Cases of Diederik Stapel, who fabricated data behind at least 30 papers (Levelt, Drenth and Noort 2012), John Darsee, who faked data behind nearly 100 publications

[1] In fact, it appears that cramped for time at least some scholars do not even carefully read the research they cite, misciting key claims (Sood and Cor 2016).

(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 8 days based on fabricated data (Service 2003; Anderson et al. 2013), are legend. So are cases of Andrew Wakefield, who published a paper linking MMR vaccine to autism using fabricated data (Wakefield et al. 1998; Deer 2011; Godlee, Smith and Marcovitch 2011), and recently Michael Lacour, who published a paper in *Science* based on fabricated data (Broockman, Kalla and Aronow 2015; McNutt 2015).[2] Each of these cases was framed as an example of misconduct by a bad actor, the subtext often being that bad actors are exceptions than the rule.

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, Anderson and De Vries (2005) (see also Titus, Wells and Rhoades (2008)). Another study found that nearly 34% of the respondents in past surveys had admitted to engaging in questionable research practices Fanelli (2009).

The other likely reason behind trust in peer reviewed research is the fact that the rate of retractions is 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 (Cokol et al. 2007). In all likelihood, however, the true rate of serious errors in manuscripts is manifolds the rate at which the errors are publicized. 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 stuffing non-significant results in the file-drawer(Franco, Malhotra and Simonovits 2014), conducting specification searches, and other more fundamental concerns like low power, which reduces the likelihood that a nominally statistically significant finding actually reflects a true effect (Button et al. 2013; Ioannidis 2005). All in all, while the belief that most research that is produced is reliable

[2]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 10 papers and numerous grant applications.

is very likely unfounded, it likely explains why scientists approvingly cite erroneous research.

Thirdly, incentives to cite carefully are mostly absent. More often than not, the only thing researchers are ever knocked on when it comes to citations is failing to cite someone or missing the journal's formatting requirements. Citing incorrectly or citing bad research flatteringly generally attracts little opprobrium.

Fourthly, when researchers are searching for relevant research, there are no tools that reliably alert researchers about errors in research. Google Scholar, for instance, does not flag if an article has been retracted, much less flag articles that have found serious problems with the article.

Lastly, often times, researchers rely on old reference databases sitting on their computer to cite new research. For instance, Davis (2012) finds that personal Mendeley libraries contained 1,340 retracted articles. And researchers likely don't check if these databases contain articles that have since been retracted because of the reasons we discuss above—chances are low. All in all, there are a lot of reasons to suspect that scientists would cite erroneous research, even where errors have been publicized via publication of a retraction notice, or another article that notes the problems.

Data and Research Design

To investigate the extent to which publicizing serious errors in research reduce citations to erroneous research, we investigate impact of publicity on citations to both, retracted articles, and articles in which a potentially serious error has been found. We include articles with serious errors that have not been retracted because retracted articles are a small subset of articles with serious errors. The bar for retraction is generally fraud, not improper statistical methods or incorrect inferences. And approving citations to the latter are a far larger problem. To study the question, we assembled a large novel dataset of retracted articles, and exploited data from an article that finds a potentially serious error in articles published in prominent scientific journals like *Science*.

To create a database of retracted articles that spans across disciplines, we used Web of Science (WoS) (Reuters 2012). WoS indexes articles from over 12,000 international journals and 148,000 conferences (Yong-Hak 2013). WoS includes key citation indices—*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.[3]

To build a database of retracted articles, we started by creating a list of retraction notices. To do that, we searched WoS for titles containing the phrase "retraction of." This yielded more than 14,000 records. Using the "corrections" filter in WoS, we filtered the list to a final set of 4,085 retraction notices.

Next, we used the list of retraction notices to search the WoS for information about the articles that were retracted. Retraction notices did not contain consistent titles to facilitate a direct search of the original articles. However, 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 sources of information along with the name of the publication to search for the original articles. This resulted in 3,776 records. We couldn't locate the remaining 309 retracted articles.

Some of the 3,776 articles, however, were false positives. Using the year of publication, list of authors, and title of publication sometimes resulted in multiple hits because the same set of authors had multiple publications in the same year in the same journal! And in some cases the people listed as authors in the retraction notice were instead editors of the journal. To flag potential false positives, we devised a set of rules. If the list of authors of the retracted article

[3]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/soodoku/propagation_of_error/data/wos/.

didn't match the list of authors for the relevant retraction notice record, we flagged the record as a potential false positive. Similarly, if the title of the retracted article didn't contain the words "retracted" or "retraction," we flagged it as a false positive. (It is standard practice for titles of original articles to be revised to indicate the article has been retracted.) Finally, we parsed the titles of the retracted notices to extract the volume and page number of the original article. Cases where volume and page number didn't match were flagged as potential false positives.

To separate false positives from true positives, we crafted three rules based on the data. Firstly, if the authors were the same, and the title contained the words "retracted" or "retraction," we coded the record as a true positive. Secondly, if the authors, volume numbers, and page numbers matched but the title did not contain the words "retracted" or "retraction," we again assumed it to be a true positive. However, if the title contained the words "retraction" or "retracted," and the volume and page number matched but authors didn't, we coded it as a false positive. Remaining potential false positives were examined manually to determine whether the record was a true false positive. This process resulted in a set of 3,359 articles. This serves as our final dataset.

Of the final list of 3,359 retracted articles, published between 1974 and 2016, only 3,096 were ever cited. These 3,096 retracted articles had received a total of 82,238 citations by August, 2016. The retraction notifications of these retracted articles, however, had only been cited a total of 2,435 times by the same time.

Our second dataset comes from articles that mistake difference between a statistical significant and statistically insignificant result as evidence that the difference is statistically significant (Nieuwenhuis, Forstmann and Wagenmakers 2011). (For an explanation of why this is problematic, see Gelman and Stern (2006).) In total, Nieuwenhuis, Forstmann and Wagenmakers (2011) analyzed 170 articles published in *Nature*, *Science*, *Neuron*, and *Journal of Neuroscience* between 2009 and 2010. They found that roughly half of the 170 articles had made this mistake. We got the list of articles that were analyzed from Nieuwenhuis, Forstmann and Wagenmakers (2011) and used WoS to download citations to all the 170 articles.

We added to the two datasets in two ways. Firstly, to measure whether citations were approving or not after the error had been publicized, we coded citations to the erroneous article in a random set of articles published after the error had been publicized. From both the WoS and the Nieuwenhuis data, we took a random sample of 100 articles citing erroneous research, and coded whether or not researchers note any concerns about the erroneous article they cite. Secondly, to understand why the articles are retracted, we coded the reasons given for retraction in a random set of 115 retraction notices. The reasons naturally fell into one of six categories: a) plagiarism, including self-plagiarism, duplication of data, words, and publishing the same or similar article in multiple journals, b) major errors, c) fraud, d) ethics violations, e) conflict over authorship or approval from other authors, and f) copyright issues.

Using the two datasets, we describe various features of citations to erroneous articles, and assess the impact of publicizing of errors on frequency of citations. We expect publication of retraction notice or an article noting a potentially serious error in an article to increase awareness about specific problematic articles. We also expect publications noting a potentially serious kind of error to increase awareness about the error. For instance, we expect publication of Nieuwenhuis, Forstmann and Wagenmakers (2011) to increase awareness of the particular error in statistical reasoning. Either pathway should lead to a decline in approving citations to the article. Though for the reasons we note above, we expect the average decline to be modest.

Lastly, we expect the decline in citations due to greater publicity about a general error to be considerably more tepid than decline due to a retraction. For greater awareness of an error in reasoning to lead to reduction in citations to articles making that error, scientists actually need to closely read the articles they cite. Some evidence suggests that many scientists do not read the articles they cite closely (Sood and Cor 2016).

To estimate the impact of publication of error on citation rates, we track citation rate a few years before and after the information about the error is made public. Given long publication cycles, and assuming the article would have been accepted for publication before the discovery of error, we test the impact of citations two and three years out. In case of Nieuwenhuis, Forstmann and Wagenmakers (2011), we also try a difference-in-difference identification strategy, exploiting the fact that roughly half of the articles published in the same journals did not have the same potentially serious error.

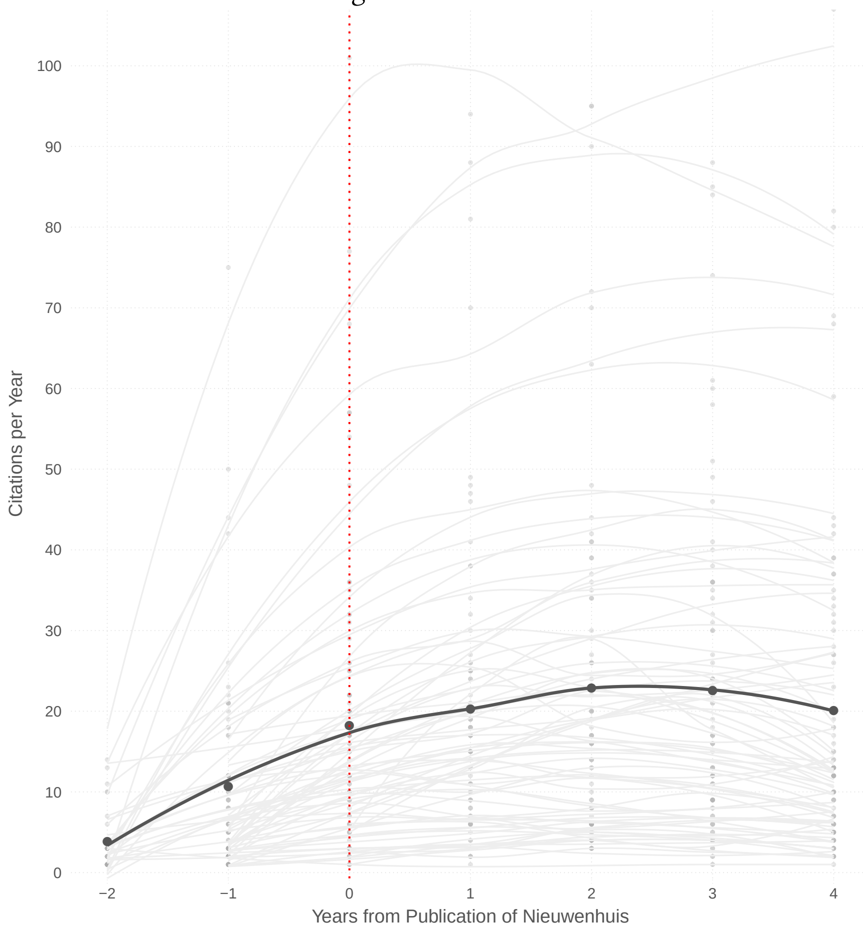
Results

We start by describing the results from Nieuwenhuis, Forstmann and Wagenmakers (2011) data, and follow it with results from the much larger retracted article data.

Prima facie evidence suggests little impact of publication of Nieuwenhuis, Forstmann and Wagenmakers (2011) on citations to articles mistaking difference between significant effect and insignificant effect as evidence for significant difference. In the two years before the publication of Nieuwenhuis, Forstmann and Wagenmakers (2011), and the year Nieuwenhuis, Forstmann and Wagenmakers (2011) was published (2011), the articles making the mistake were cited 2,267 times. Between 2012 and 2015, the articles were cited an additional 6,604 times.

Figure 1 offers a closer look. It plots 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. Two more things become clear: a) there is a skew in citation rates (skewness = 2), b) the pattern that we see in the aggregate is ecological—there is little evidence of decline in citations of any article post publication of Nieuwenhuis, Forstmann and Wagenmakers (2011). To account for the skewness, we switched means with medians. Doing so yields a pretty similar pattern with the expected intercept shift (see Figure SI 1.1). Subsetting on articles where the mistake, according to Nieuwenhuis, Forstmann and Wagenmakers (2011), has potentially serious implications for the result shows a similar pattern (see Figure SI 1.2).

Figure 1: Impact of Publication of Nieuwenhuis, Forstmann and Wagenmakers (2011) on Mean Number of Citations to Articles Containing the Error.



The plot shows total number of citations received per year by each of the papers making the mistake, and the average number of citations received per year by those articles.

But citations to erroneous research don't need to decline after the error in the research is publicized. We only expect approving citations—citations that do not note any concerns—to decline. As we note above, we coded citations to erroneous articles in a randomly chosen set of 100 articles published after the error was publicized. We couldn't locate one of the 100 articles, leaving us with a sample of 99 articles. Of the 99 articles, 2 were false positives—the articles didn't cite erroneous research, but 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 the mistake, citing Nieuwenhuis, Forstmann and Wagenmakers (2011) for support.

To more formally explore change in citation rate as a consequence of publication of Nieuwenhuis, Forstmann and Wagenmakers (2011), we regress citations per year on a dummy for the year Nieuwenhuis, Forstmann and Wagenmakers (2011) was published, a linear time trend, and fixed effect for 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 Nieuwenhuis, Forstmann and Wagenmakers (2011) is published (see Table 1).

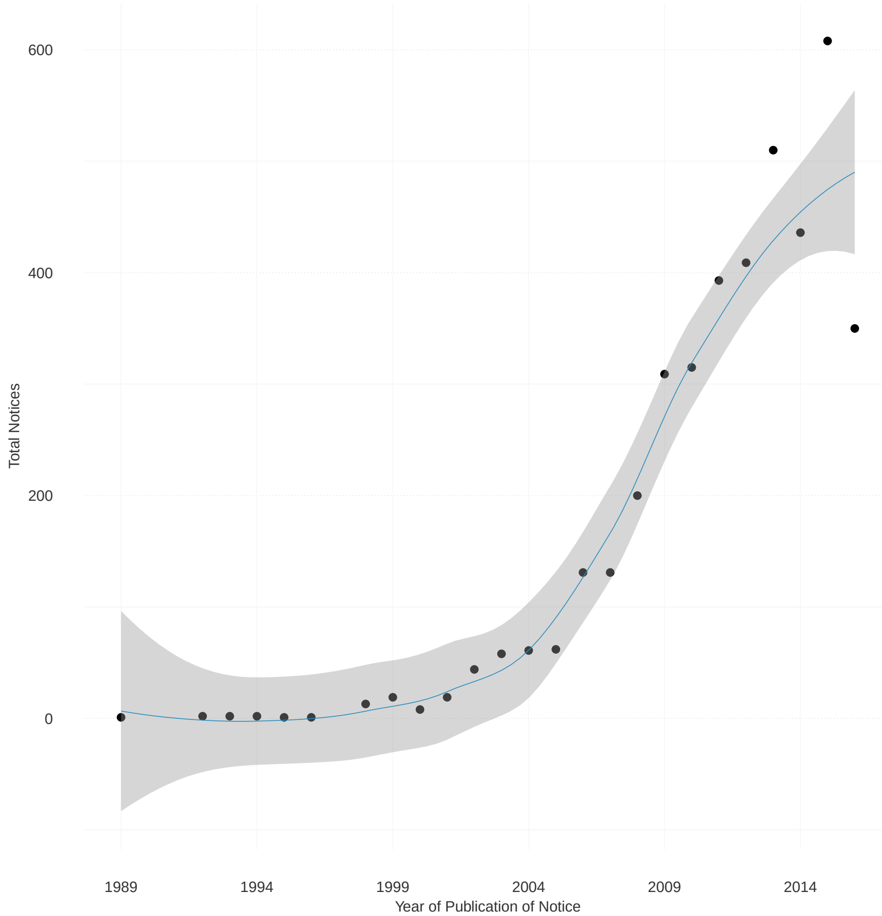
Table 1: Impact of Publication of Nieuwenhuis on Citations to Articles with Potentially Serious Errors

	Dependent variable:
	Citation Count
Transition Date	3.808**
	(1.703)
Time	2.024***
	(0.443)
Constant	20.758***
	(4.019)
Observations	487
R^2	0.784
Adjusted R^2	0.743
Note:	*p<0.1; **p<0.05; ***p<0.01

Next, we leverage the novel retracted article dataset to shed light on the process of retraction. We start with descriptive data that shed some light on a few important features of scientific retractions—number of retractions over time, why articles are retracted, which research fields tend to have the most retractions, and average time till retraction.

Over the last thirty or so years, the number of retractions have increased sharply (see Figure 2). First retraction notice that we have in our database is from 1989. That year and decade after it, the number of retraction notices being published each year never crossed 20. Since then, there has been a sharp and accelerating increase in retraction notices per year. Between 2001, when 19 retraction notices were published, and 2015, last year for which we have complete data, there was a more than 30 fold increase. There were a total of 608 retraction notices in 2015. The pattern that we find is consistent with results from Steen, Casadevall and Fang (2013), who also find a rapid increase in retractions over time. A good chunk of the increase is likely explained by greater production of research over time, but the particularly sharp increase in the last 15 years suggests that greater editorial awareness and new tools that reduced cost of detection of issues like plagiarism likely played a major role.

Figure 2: Retraction Notices Per Year

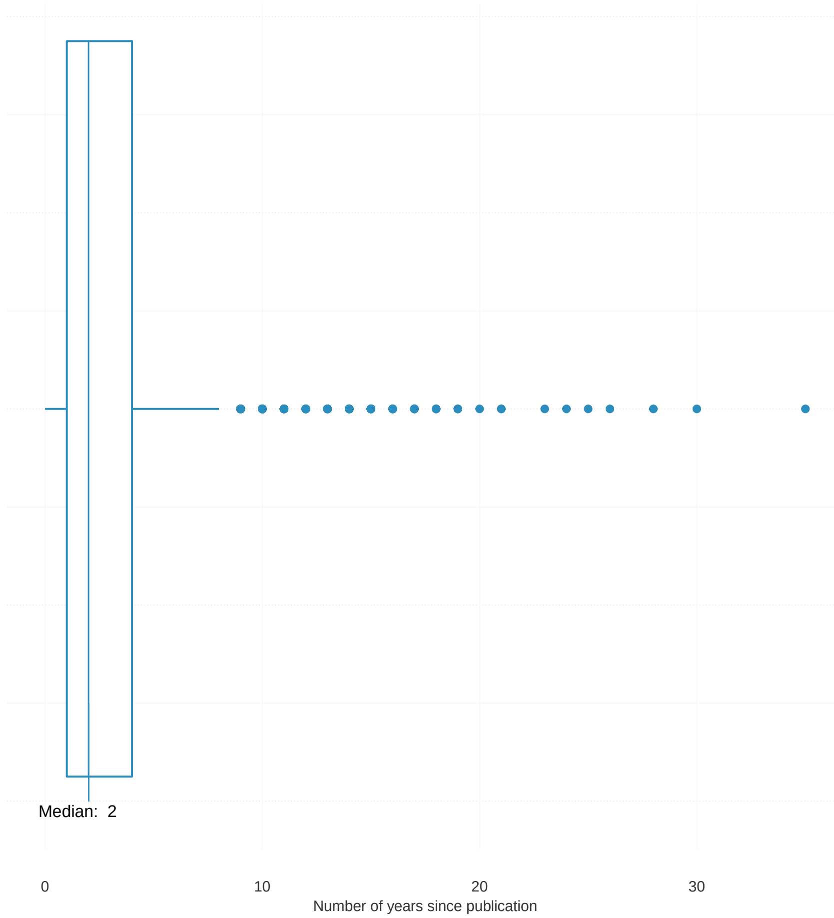


Of the 115 notices, about 50 articles were retracted due at least in part to plagiarism. (We define plagiarism to include self-plagiarism, duplication of data, words, and publishing the same or similar article in multiple journals.) Major errors or fraud contributed to another 58 retractions, with fraud alone accounting for 27 retractions. Ethics violations (2), conflict over authorship, or approval from other authors (5), copyright (1) contributed to the rest. 50.4% rate of retraction due to fraud that we find is not too similar for some 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, Coble and Anderson 2011) (see also Steen (2010)). Plagiarism also continues to be a significant problem. A study of biomedical literature found close to 3,000 publications each year that are "highly similar to citations in previously published manuscripts" (Garner 2011).

Next, we tallied which research areas of the retracted articles. We used advanced search and research area filter in WoS to estimate the number of articles written in English in each of the major research areas that the WoS carried. As of November 1st, 2016, WoS had 363,363 articles in Social Sciences, 287,379 articles in Life Sciences & Biomedicine, 23,532 articles in Physical Sciences, 2,220,516 articles in Technology, and 117,376 articles in Arts & Humanities. And while WoS carries far more Social Science articles than Life Science and Physical Science articles, there were far more retractions in the latter two categories than the former. And perhaps unsurprisingly, there were no arts and humanities retractions. (For a complete list of retractions by research field, see SI 2.)

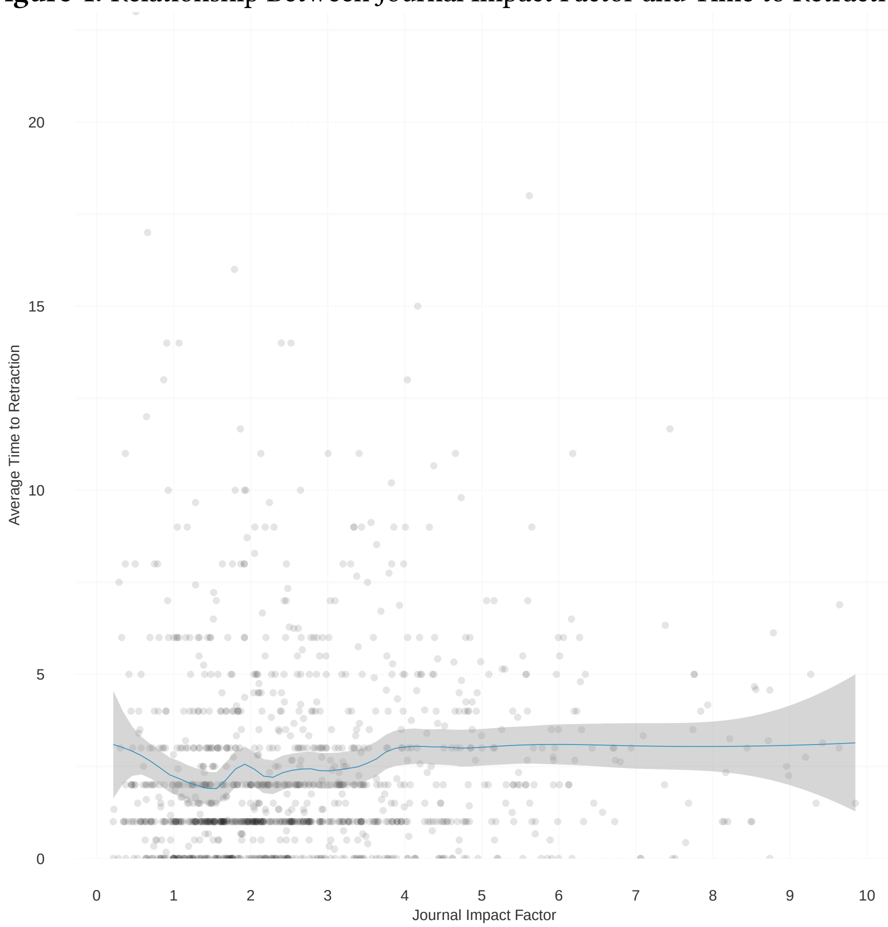
In total, the articles that were retracted were cited 41,347 before they were retracted by November, 2016. On average, it took 2.94 years for the article to be retracted; the median time was 2 years (see Figure 3.) It took more than 25% of the articles 4 or more years. And it took 35 years to retract one article. These numbers compare favorably to a study on time to retraction in the PubMed corpus. Steen, Casadevall and Fang (2013) found that the average time to retraction was nearly 3 years on average, with time to retraction declining over time — from nearly 4 years for articles published in or before 2002 to just over 2 years for articles published after.

Figure 3: Time to Retraction



Next, we estimated the relationship between journal impact factor and average time to retraction on the hunch that prominent journals would attract greater readership, which in turn would more quickly flag problematic research. Surprisingly, there is no relationship between journal impact factor and average time to retraction—flawed articles in low ranked journals are retracted as quickly as flawed articles in higher ranked journals (see figure 4).

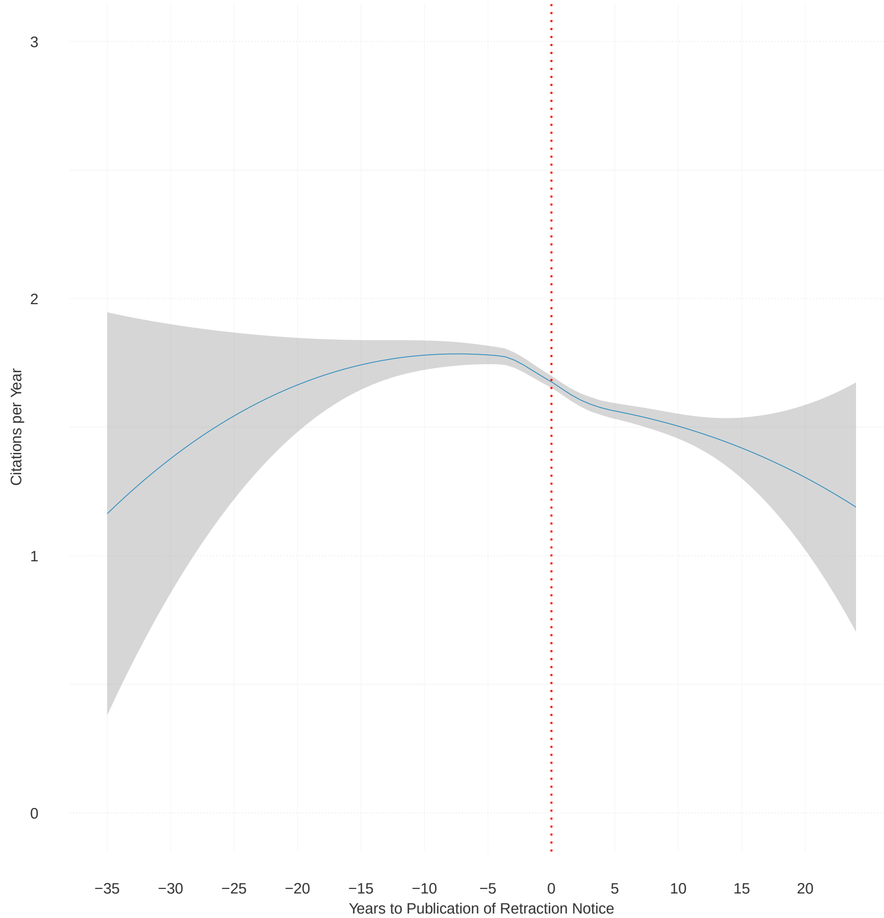
Figure 4: Relationship Between Journal Impact Factor and Time to Retraction



Next, we assessed the impact of publication of retraction on citation. As we note above, retracted articles were cited 41,347 before they were retracted. The retracted articles were cited another 40,872 between the time they were retracted and August, 2016. As we note above, retraction notifications were cited a total of 2,435 times. Assuming citation to retraction notification means not citing the retracted article approvingly, one crude estimate of the total citations that took retracted studies' results to be valid is 38,437. Thus, on average, the 3,268 retracted articles received an additional 11.76 citations after the retraction notice was published and before August, 2016.

To look more carefully at the impact of publication of citation on frequency of citation, we plotted a LOESS over total citations to an article per year. Figure 5 shows a small downturn that coincides with retracted, but followed by a plateauing.

Figure 5: Impact of Publication of Retractions on Citations to Articles



Discussion

Do citations to research with serious errors drop after the errors are publicized? Or does research containing serious errors continue to be cited apace, propagating the error apace? Data suggest that publicizing serious errors via public retractions or publication of research highlighting the problem at prominent venues leads to, at best, a modest decline in citations. Retracted articles continue to be cited approvingly years after they have been publicly retracted.

Our results echo conclusions reached by other research on the topic based on much smaller corpora. Kochan and Budd (1992), for instance, found that John Darsee's papers continued to be approvingly cited even after a considerable time after retraction, and even though the case had generated much publicity. Another study found that although retraction reduces subsequent citation compared with a control group, retracted papers were often cited to support claims (Pfeifer and Snodgrass 1990). Similarly, Budd, Sievert and Schultz (1998) used Medline to identify retracted articles and found that many retracted articles were still being cited as valid.[4]

Citations to flawed research are likely consequential. Such citations very likely affect people's beliefs about the preponderance of evidence on a point. These 'mistakes'—citing flawed research when flaws have been made public—are also avoidable. Assuming that researchers do not knowingly approvingly cite retracted articles, the data imply that discovery of errors even when public retraction notices are issued is still a problem.

To ameliorate the problem, we need to improve access to information about problems in research. One way to improve access to information about problems is to build tools that provide the information as part of existing research discovery and production processes. For instance, altering interfaces of heavily used portals such as Google Scholar, JSTOR, journal publishers' sites etc. so that they thread reproduction attempts, retractions, and other research that directly bears on the evidence presented in an article along with the article are liable to be effective. Rather than effect change in multiple interfaces, which requires coordination with multiple strategic actors, however, a better strategy may be to create a browser plug-in that highlights problematic articles listed on a web page. Providing such a tool to editors or copy editors at academic publishers may also help ameliorate the problem. Flagging problems during the scientific discovery process, however, is clearly better than flagging them during the production process. Flagging during discovery likely preempts the temptation to engage in post hoc rationalization. Alternately, one could build tools that automatically create pull requests to personal bibliography libraries posted

[4]Other research has focused on analyzing the impact of retraction on citations to other work by authors of the retracted research, finding that citations to other research declines when a publication is retracted (Lu et al. 2013; Azoulay, Bonatti and Krieger 2015).

on open publication platforms like GitHub. Lastly, while our study only tallies research that cites known flawed research, it is quite likely that the effect of flawed research extends to studies that cite studies that approvingly cite flawed research (and thereon). And any modifications to the interface should extend to papers that cite flawed research so that people citing them in turn are also warned.

Egregious errors like approving citations to flawed research after the flaws have been made public serve to highlight larger problems with how science is practiced. Scholars do not appear to carefully vet research they cite. In fact, data suggest that scholars do not even always carefully read the research they cite, misciting key claims (Sood and Cor 2016). To improve reliability of scientific production, besides innovating on better tools, we may also need to also penalize research that makes such errors.

References

Anderson, Melissa S, Marta A Shaw, Nicholas H Steneck, Erin Konkle and Takehito Kamata. 2013. Research integrity and misconduct in the academic profession. In *Higher education: handbook of theory and research*. Springer pp. 217–261.

Azoulay, Pierre, Alessandro Bonatti and Joshua L Krieger. 2015. The career effects of scandal: Evidence from scientific retractions. Technical report National Bureau of Economic Research.

Basu, Paroma. 2006. "Where are they now?" *Nature medicine* 12(5):492–493.

Broockman, David, Joshua Kalla and Peter Aronow. 2015. "Irregularities in LaCour (2014)." *Work. pap., Stanford Univ. http://stanford. edu/ dbroock/broockman_kalla_aronow_lg_irregularities. pdf* .

Budd, John M, MaryEllen Sievert and Tom R Schultz. 1998. "Phenomena of retraction: reasons for retraction and citations to the publications." *JAMA* 280(3):296–297.

Budd, John M, Zach C Coble and Katherine M Anderson. 2011. Retracted publications in biomedicine: Cause for concern. In *Association of College and Research Libraries Conference*. pp. 390–5.

Button, Katherine S, John PA Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma SJ Robinson and Marcus R Munafò. 2013. "Power failure: why small sample size undermines the reliability of neuroscience." *Nature Reviews Neuroscience* 14(5):365–376.

Cokol, Murat, Ivan Iossifov, Raul Rodriguez-Esteban and Andrey Rzhetsky. 2007. "How many scientific papers should be retracted?" *EMBO reports* 8(5):422–423.

Davis, Philip 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(3):184.

Deer, Brian. 2011. "How the case against the MMR vaccine was fixed." *BMJ* 342:c5347.

Fanelli, Daniele. 2009. "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data." *PloS one* 4(5):e5738.

Franco, Annie, Neil Malhotra and Gabor Simonovits. 2014. "Publication bias in the social sciences: Unlocking the file drawer." *Science* 345(6203):1502–1505.

Garner, Harold R. 2011. Combating unethical publications with plagiarism detection services. In *Urologic Oncology: Seminars and Original Investigations.* Vol. 29 Elsevier pp. 95–99.

Gelman, Andrew and Hal Stern. 2006. "The difference between "significant" and "not significant" is not itself statistically significant." *The American Statistician* 60(4):328–331.

Godlee, Fiona, Jane Smith and Harvey Marcovitch. 2011. "Wakefield's article linking MMR vaccine and autism was fraudulent." *BMJ* 342:c7452.

Ioannidis, John PA. 2005. "Why most published research findings are false." *PLoS Med* 2(8):e124.

Kochan, Carol Ann and John M Budd. 1992. "The persistence of fraud in the literature: the Darsee case." *Journal of the American Society for Information Science* 43(7):488.

Levelt, Willem JM, PJD Drenth and E Noort. 2012. "Flawed science: The fraudulent research practices of social psychologist Diederik Stapel.".

Lu, Susan F, Ginger Zhe Jin, Brian Uzzi and Benjamin F Jones. 2013. "The retraction penalty: evidence from the Web of Science.".

Martinson, Brian C, Melissa S Anderson and Raymond De Vries. 2005. "Scientists behaving badly." *Nature* 435(7043):737–738.

McNutt, Marcia. 2015. "Editorial retraction." *Science* p. aaa6638.

Nieuwenhuis, Sander, Birte U Forstmann and Eric-Jan Wagenmakers. 2011. "Erroneous analyses of interactions in neuroscience: a problem of significance." *Nature neuroscience* 14(9):1105–1107.

Pfeifer, Mark P and Gwendolyn L Snodgrass. 1990. "The continued use of retracted, invalid scientific literature." *JAMA* 263(10):1420–1423.

Reuters, Thomson. 2012. "Web of Science.".

Service, Robert F. 2003. "Scientific misconduct. More of Bell Labs physicist's papers retracted." *Science (New York, NY)* 299(5603):31.

Sood, Gaurav and Ken Cor. 2016. "A Citation for Miscitation: Estimate of the Frequency of Miscitations in Academic Articles.".

Steen, R Grant. 2010. "Retractions in the scientific literature: is the incidence of research fraud increasing?" *Journal of medical ethics* pp. jme–2010.

Steen, R Grant, Arturo Casadevall and Ferric C Fang. 2013. "Why has the number of scientific retractions increased?" *PLoS One* 8(7):e68397.

Stewart, Walter W and Ned Feder. 1987. "The integrity of the scientific literature." *Nature* 325:207–14.

Titus, Sandra L, James A Wells and Lawrence J Rhoades. 2008. "Repairing research integrity." *Nature* 453(7198):980–982.

Wakefield, Andrew J, Simon H Murch, Andrew Anthony, John Linnell, DM Casson, Mohsin Malik, Mark Berelowitz, Amar P Dhillon, Michael A Thomson, Peter Harvey et al. 1998. "RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children." *The Lancet* 351(9103):637–641.

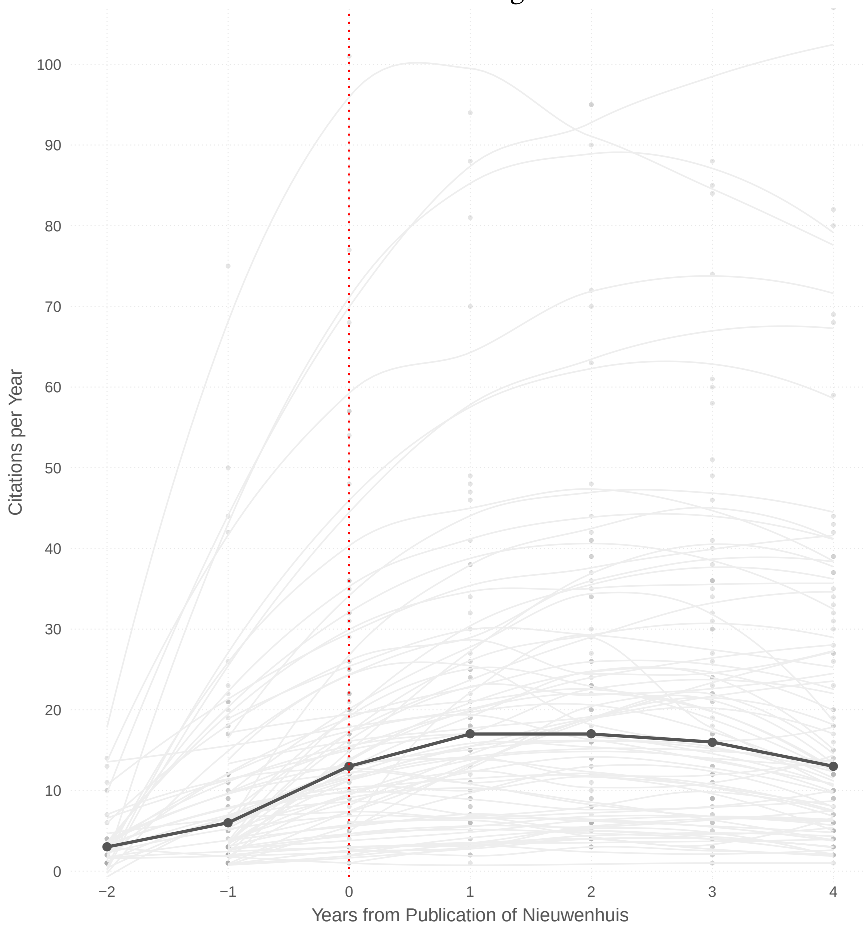
Wallis, Claudia. 1983. "Fraud in a Harvard lab." *Time* 121(9):49–49.

Yong-Hak, Jo. 2013. "Web of Science." *Thomson Reuters* .

Supporting Information

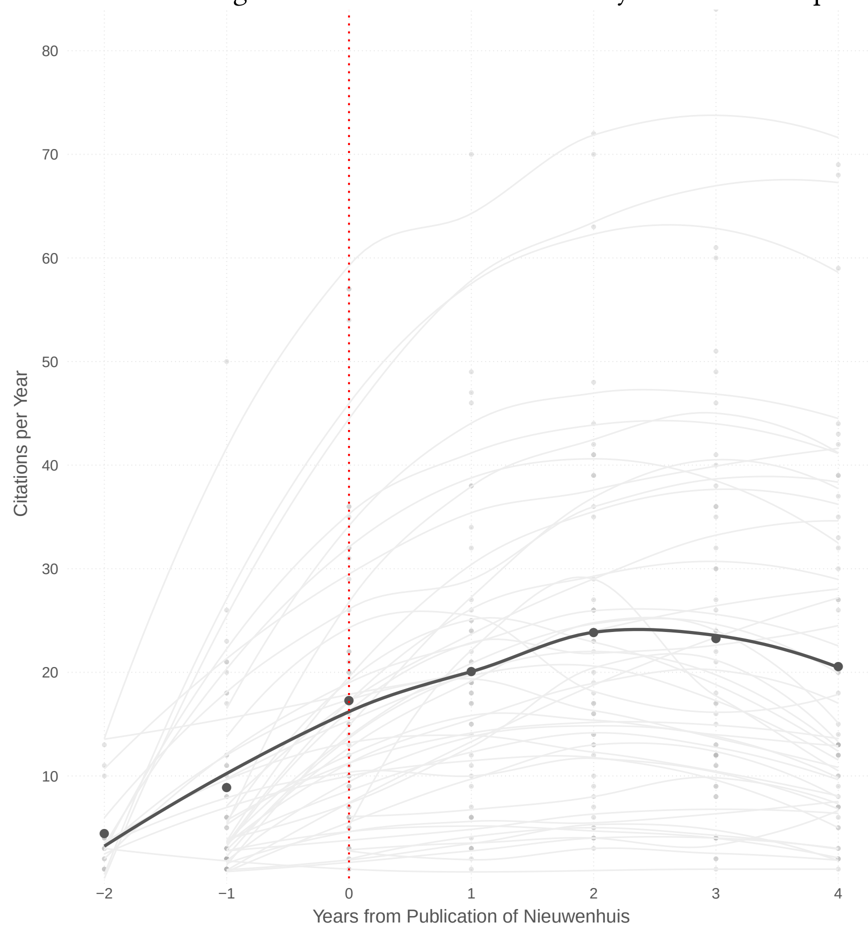
SI 1

Figure SI 1.1: Impact of Publication of Nieuwenhuis, Forstmann and Wagenmakers (2011) on Median Number of Citations to Articles Containing the Error.



The plot shows 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 articles.

Figure SI 1.2: Impact of Publication of Nieuwenhuis, Forstmann and Wagenmakers (2011) on Citations to Articles Containing the Error Which has Potentially Serious Consequences for the Results.



The plot shows total number of citations received per year by each of the papers making the mistake with 'potentially serious' consequences and the average number of citations received per year by those articles.

SI 2 Retraction Notices by Web of Science Field

The fields are decided by Web of Science.

Table SI 2.2: Retraction Notices By Field

Field	Number of Notices	Percentage of Total
BIOCHEMISTRY MOLECULAR BIOLOGY	431	10.44
CELL BIOLOGY	306	7.41
SCIENCE TECHNOLOGY OTHER TOPICS	283	6.86
ENGINEERING	282	6.83
CHEMISTRY	282	6.83
ONCOLOGY	256	6.20
NEUROSCIENCES NEUROLOGY	198	4.80
PHYSICS	188	4.55
PHARMACOLOGY PHARMACY	187	4.53
MATERIALS SCIENCE	179	4.34
ANESTHESIOLOGY	165	4.00
SURGERY	158	3.83
IMMUNOLOGY	155	3.75
RESEARCH EXPERIMENTAL MEDICINE	154	3.73
CARDIOVASCULAR SYSTEM CARDIOLOGY	143	3.46
GENERAL INTERNAL MEDICINE	130	3.15
ENVIRONMENTAL SCIENCES ECOLOGY	114	2.76
PSYCHOLOGY	98	2.37
MECHANICS	90	2.18
HEMATOLOGY	90	2.18
ENDOCRINOLOGY METABOLISM	90	2.18
BIOTECHNOLOGY APPLIED MICROBIOLOGY	87	2.11
BUSINESS ECONOMICS	83	2.01
MATHEMATICS	82	1.99
GENETICS HEREDITY	79	1.91
BIOPHYSICS	64	1.55
RESPIRATORY SYSTEM	63	1.53
MICROBIOLOGY	62	1.50
LIFE SCIENCES BIOMEDICINE OTHER TOPICS	60	1.45
CRYSTALLOGRAPHY	56	1.36
ENERGY FUELS	54	1.31
ACOUSTICS	53	1.28
GASTROENTEROLOGY HEPATOLOGY	51	1.24
COMPUTER SCIENCE	49	1.19
PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH	48	1.16
PLANT SCIENCES	46	1.11
PHYSIOLOGY	44	1.07
OBSTETRICS GYNECOLOGY	44	1.07
ORTHOPEDICS	43	1.04
METALLURGY METALLURGICAL ENGINEERING	42	1.02
PATHOLOGY	39	0.94
FOOD SCIENCE TECHNOLOGY	39	0.94
UROLOGY NEPHROLOGY	38	0.92
AGRICULTURE	37	0.90
NUTRITION DIETETICS	36	0.87
INFECTIOUS DISEASES	34	0.82

VIROLOGY	30	0.73
PSYCHIATRY	30	0.73
TOXICOLOGY	29	0.70
DENTISTRY ORAL SURGERY MEDICINE	28	0.68
TRANSPLANTATION	27	0.65
PEDIATRICS	27	0.65
RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING	26	0.63
OPHTHALMOLOGY	24	0.58
POLYMER SCIENCE	23	0.56
WATER RESOURCES	21	0.51
REPRODUCTIVE BIOLOGY	20	0.48
OPTICS	20	0.48
GEOLOGY	20	0.48
INSTRUMENTS INSTRUMENTATION	19	0.46
DEVELOPMENTAL BIOLOGY	19	0.46
RHEUMATOLOGY	18	0.44
THERMODYNAMICS	17	0.41
NURSING	16	0.39
LITERATURE	16	0.39
SPORT SCIENCES	14	0.34
OPERATIONS RESEARCH MANAGEMENT SCIENCE	14	0.34
ELECTROCHEMISTRY	14	0.34
EDUCATION EDUCATIONAL RESEARCH	14	0.34
DERMATOLOGY	14	0.34
VETERINARY SCIENCES	13	0.32
PUBLIC ADMINISTRATION	12	0.29
NUCLEAR SCIENCE TECHNOLOGY	12	0.29
METEOROLOGY ATMOSPHERIC SCIENCES	12	0.29
HEALTH CARE SCIENCES SERVICES	12	0.29
ZOOLOGY	11	0.27
REHABILITATION	11	0.27
MEDICAL LABORATORY TECHNOLOGY	11	0.27
EVOLUTIONARY BIOLOGY	11	0.27
EMERGENCY MEDICINE	10	0.24
OTORHINOLARYNGOLOGY	9	0.22
MATHEMATICAL COMPUTATIONAL BIOLOGY	9	0.22
GOVERNMENT LAW	9	0.22
GERIATRICS GERONTOLOGY	9	0.22
SOCIOLOGY	8	0.19
PHYSICAL GEOGRAPHY	8	0.19
INTEGRATIVE COMPLEMENTARY MEDICINE	8	0.19
GEOGRAPHY	8	0.19
BEHAVIORAL SCIENCES	8	0.19
AUTOMATION CONTROL SYSTEMS	8	0.19
ANATOMY MORPHOLOGY	8	0.19
ALLERGY	8	0.19
TELECOMMUNICATIONS	7	0.17
MARINE FRESHWATER BIOLOGY	7	0.17
ASTRONOMY ASTROPHYSICS	7	0.17
ANTHROPOLOGY	7	0.17
SPECTROSCOPY	6	0.14
SOCIAL SCIENCES OTHER TOPICS	6	0.14
INTERNATIONAL RELATIONS	6	0.14

ENTOMOLOGY	6	0.14