10

Retraction Watch: What We've Learned and How Metrics Play a Role

Ivan Oransky

Adam Marcus and I founded Retraction Watch (http://retractionwatch.com) in August 2010 for two reasons: As longtime journalists, we often found that retraction notices were opaque. And sometimes opacity was the best you could hope for; often, notices were misleading or even wrong. We also found that there were great stories behind retractions.

We have our own metrics at Retraction Watch, mostly just having to do with traffic to the site each month; we now have, on average, 150,000 unique visitors, and half a million page views. (However, we are not beholden to these metrics, as our revenue does not depend on advertising; we have at various times had generous funding from three foundations, and other income streams including freelance writing fees.) In terms of more traditional metrics, I can say we have been cited in the literature more than a hundred times. That means that if a blog could have an H index, we would have a good one. And it does not hurt when we talk to funders about the impact we are having on publishing practices and transparency.

Retraction Watch posts often begin with a tip—mostly a notice of retraction. But we also receive long emails from frustrated researchers, who have been laboring to correct a perceived wrong for months, if not years. We empathize and sympathize with their frustration—it is incredibly hard to get papers retracted from the literature, or even corrected or noted in some way.

As an illustration, take a piece by nutrition researcher David Allison and colleagues that appeared in *Nature* (Allison et al., 2016). They scanned the nutrition literature and found more than two dozen papers that they thought were deeply problematic. And they kept a pretty high bar. You can judge for yourselves, but if you look at the kinds of problems they were looking at in these papers, it was pretty clear something needed to be done. In a few cases, the journals retracted the paper, or published a letter from Allison and his team critiquing the findings, but in many cases the

journals did nothing. It was very, very frustrating. We wrote a commentary on this eye-opening article (Oransky and Marcus, 2016) and a Q&A with the authors as well (McCook, 2016a). And although retraction is not always the best way to correct the scientific literature—corrections and correspondence, for example, may serve the purpose—the nuclear option is sometimes necessary. But it is very, very difficult to get a paper retracted. That is for a number of reasons, but one is clearly the stigma attached to retractions—although I will present some surprising findings on that in the conclusion of this chapter—and the fact that publishing papers in certain journals is the only way to earn grants, tenure, and promotions. They are, in a word, the metrics. Researchers will do anything to publish papers in some journals, including even creating fake authors (Marcus and Oransky, 2016). The opposite side of that coin is that many authors are very reluctant to admit to flaws in their work, and, by extension, to retract. I would put Allison and his colleagues in a "still relatively patient and professional" category in terms of their approach to trying to correct the record. But there are others who let their exasperation show. Some of our frequent commenters fall into that category, and they often find it difficult to maintain their equanimity. It can be very frustrating to try and correct the literature. Often well-meaning researchers who care about the public record use their free time to correct it, but it does not go well and they are vilified. To put it another way, whistle-blowers often fare far worse than those they are accusing of misconduct—and most problematic papers do not even earn an Expression of Concern, let alone a retraction. In short, the most common outcome for those who commit fraud is: a long career.

There is a really robust source of these conversations about the literature: PubPeer (Barbour and Stell, this volume, chapter 11). The site became a great resource for us, because we had a place to refer all the readers who were sending us allegations that we are not equipped—in terms of time, resources, and expertise—to properly vet and write about. PubPeer comments are a great resource for us, but so are the comments on our own site, in response to particular posts. There is much debate over comments and whether or not commenters should be allowed to remain anonymous (Blatt, 2015). We have chosen to allow commenters to post anonymously, because of the hierarchy of science and the dangers of speaking out against the powerful. But we moderate each comment heavily, and won't post any allegations that cannot be supported by evidence.

Here is an example of why we believe in the importance of anonymous comments. In 2014, we learned of a series of retractions from diabetes researcher Cory Toth, who used to be at the University of Calgary. We

reported on them, and he actually said some very interesting things. "I am significantly apologetic, remorseful, and embarrassed that this occurred under my watch," he told us. "Please know that I will not be publishing in the world of science in the future" (Marcus, 2014). The university had done an investigation, and they found misconduct, so he left.

Once we posted about four retractions, things became even more interesting. Readers started leaving comments that said, basically, "Oh, what about this paper that wasn't retracted? You should look at this other figure. It looks a little bit odd. Something else here, there, and everywhere." The university—unbeknownst to us—saw these comments and followed up on them. They reopened their investigation, and he ended up with five more retractions. And when Margaret Munro, then a reporter for Postmedia, called the university, they told her it was because of the Retraction Watch comments (Munro, 2014).

That story, and PubPeer, are important reminders that we are part of an ecosystem. What happens on PubPeer can result in a retraction (Keith, 2015). I can point to dozens of cases we have covered where the allegations first appeared in PubPeer. But it is not just us and PubPeer. There are many people out there focusing on similar issues. Some, like Science -Fraud.org, have been forced to shutter because of legal threats (Brookes, this volume, chapter 13; Oransky, 2013). And PubPeer faced legal action brought by a scientist who lost a job offer at least in part due to comments on the site. The scientist wanted PubPeer to unmask a commenter, but with the help of the American Civil Liberties Union, earned a judge's ruling that allowed them to refuse (McCook, 2016b).

While we face those kinds of threats, too, we have yet to be sued, which we attribute to having worked in journalism for more than a decade and having developed a clear sense of libel laws and how they apply to our efforts. Also, because we are journalists, rather than practicing scientists, we can write critically of researchers, publishers, and institutions without fearing reprisals from those in power that would damage our own careers. There are others in our ecosystem, like Ben Goldacre, who are going from strength to strength. Many of us have been funded by the Arnold Foundation, whose support of this area should be acknowledged. It is very important if you get involved in this work to think of yourselves as part of an ecosystem.

To put it another way, there are lots of people out there who are as obsessive about their niches as Adam and I are about retractions. We are eager to host those fellow obsessives, in guest posts—such as a post about the Collaboration Score from its creator, Sarah Greene (Greene, 2016).

What did we learn in our first five years? One clear trend is that the retraction rate is on the rise. The number of retractions has also gone up, not surprisingly, as there are more papers. From 2001 to 2010, as very ably demonstrated in *Nature* by Richard Van Noorden, the number of retractions increased from about forty a year to four hundred. The number of papers only went up forty-four percent (Van Noorden, 2011). The number of retractions went up again to somewhere between five hundred and six hundred, and in fiscal year 2015, it was close to seven hundred, according to PubMed. And there are more retractions that are not captured by typical databases. (Keep in mind that despite the increase, the rate of retraction remains quite small—less than 0.05 percent of all papers.) Our database of retractions, made available in October 2018 at retractiondatabase.org years after the conference from which this chapter originated, demonstrated that those trends continued but that the rate of retraction may be leveling off (Brainard and You, 2018).

Does this mean that fraud is on the increase? We do not know, but we are certainly better at catching it. The introduction of plagiarism detection software changed a lot, as it enabled anyone to scan the millions of online papers. It is like the introduction of a new screening test: you would expect to see an increase in diagnoses once you start screening. At the same time, we have some evidence that misconduct of at least one kind—inappropriate image manipulation—is on the rise, thanks to a painstaking effort by Elisabeth Bik to scrutinize the Western blots in more than twenty thousand papers. Bik and her colleagues found that one in twenty-five papers contained problematic images—and that the rate of such issues had grown dramatically since 2000 (Bik, Casadevall, and Fang, 2016).

These are some of the reasons papers are retracted, in a fairly random order:

- **Plagiarism**: This is responsible for about ten percent of retractions.
- **Duplication**: You cannot really plagiarize yourself, but you can duplicate your own work. That is probably about fifteen percent of retractions.
- **Image manipulation**: Photoshopped photos of Western blots are a common reason for retraction, and a frequent source of comments on PubPeer.
- **Faked data**: Diederik Stapel may be the most well-known recent case of this, but it is not uncommon among reasons for retraction.
- "**For legal reasons**": This one is sort of strange to us because it suggests that all the notices that do not include that term are illegal.

- **Fake peer reviews**: This is a trend we have been reporting on since 2012. It is relevant to a discussion of metrics because you can manage to have your paper accepted if you do your own peer review, taking advantage of the way editorial management systems are set up (Ferguson, Marcus, and Oransky, 2014). More than six hundred papers have now been retracted for that reason. And the fact that researchers do this in response to "publish or perish" pressures suggests a direct link to metrics.
- **Honest error**: This is responsible for about twenty percent of retractions, and when we see authors going out of their way to correct the record, we call attention to it by adding the case to our "doing the right thing" category.
- **Publisher error**: This is not a particularly important reason for retraction, and not all that common, but sometimes publishers print the same paper two or even three times. Apparently, the only way they know to correct the record is to issue a retraction, which means an author has a retraction—and the stigma that comes with it—for no fault of his or her own.
- **Authorship issues**: These can be a real headache for journals and universities. We have seen cases in which authors appear on papers they had nothing to do with, and other cases in which rightful authors are not named.
- **Lack of reproducibility**: This is controversial. Most scientists we speak to argue against retracting a paper just because later work overturns it, but some decide to retract their own papers for this reason.

Overall, two-thirds of retractions are due to misconduct, as Fang, Steen, and Casadevall showed in 2012. Sometimes, however, it is hard to tell. Here is a notice that was typical in a particular journal until quite recently: "This article has been withdrawn by the authors." That is not very helpful, but the *Journal of Biological Chemistry*, where the "article has been withdrawn by the authors" retractions appeared, must have become tired of having us beat them up after five years, so they changed their policy and now include details (Guengerich, 2015).

Speaking of metrics, here is how they might be related to how long it takes to retract. The average is about three years (Steen, Casadevall, and Fang, 2013). To be a conspiracy theorist for a moment, it is worth noting that three years is a year longer than the amount of time citations count toward a journal's impact factor. That means if journals can drag out the process, they would not take an impact factor hit. Similarly, authors and

universities can drag out the process to make it less likely that a retraction will affect promotion or funding decisions.

And retracted papers keep being cited, often as if they had never been retracted (Budd, Coble, and Anderson 2011). As an illustration, we have a leaderboard of the top ten cited retracted papers (Retraction Watch, 2016b). Number two on the list at the time of this writing was Andrew Wakefield's infamous *Lancet* paper claiming a link between autism and vaccines (Wakefield et al., 1998). (Look, I just cited it.) And number one has had far more citations after it was retracted than before. Whether those are positive or negative citations, we do not know.

Which journals retract most? It turns out that journals with the highest impact factor—there are those metrics again–also have the highest rate of retraction. Again, we think that is mostly due to the fact that there are more eyeballs on those journals, although it is certainly possible that scientists are pushing the envelope to publish in those journals in ways that would constitute misconduct. "Meta-scientists" continue to debate the data on whether "publish or perish" plays a significant role in misconduct (Fanelli et al., 2017). And which researchers retract most? We like our leaderboards and our rankings at Retraction Watch. The top retraction holder has 183 (Retraction Watch, 2016a).

Let me leave you with an interesting finding that relies on metrics. A study—whose findings have been replicated by another group (McCook, 2015)—found that when people retract papers for fraud, you see what you would expect: their citations drop (Lu et al., 2013). In fact, citations in their whole subspecialty drop by about ten percent to fifteen percent. That is bad news, which, again, you would expect. But if you retract for honest error, however, and it is clear in the notice that the mistake is the result of honest error, you do not see that drop. So there is an example of using citations—a metric—to figure out what happens after a retraction. And it is also good news, in that good behavior is not punished.

Metrics: we just cannot get away from them.

References

Allison, David B., Andrew W. Brown, Brandon J. George, and Kathryn A. Kaiser. 2016. "Reproducibility: A Tragedy of Errors." *Nature* 530:27–29.

Bik, Elisabeth M., Arturo Casadevall, and Ferric C. Fang. 2016. "The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications." *mBio* 7(3):e00809–e0080916. doi:10.1128/mBio.00809–16.

Blatt, Michael R. 2015. "Vigilante Science." *Plant Physiology* 169(2):907–909.

Brainard, J., and J. You. 2018. "Rethinking retractions." *Science* 362(6413): 390-393.

Budd, John C., Zach C. Coble, and Katherine M. Anderson. 2011. "Retracted Publications in Biomedicine: Cause for Concern." In *Association of College and Research Libraries Conference*, pp. 390–395.

Fanelli, Daniele, Rodrigo Costas, Ferric C. Fang, Arturo Casadevall, and Elisabeth M. Bik. 2017. "Why Do Scientists Fabricate and Falsify Data? A Matched-Control Analysis of Papers Containing Problematic Image Duplications." *bioRxiv*. http://biorxiv.org/content/early/2017/04/12/126805.

Fang, Ferric C., R. Grant Steen, and Arturo Casadevall. "Misconduct Accounts for the Majority of Retracted Scientific Publications." *Proceedings of the National Academy of Sciences* 109(42):17028–17033.

Ferguson, Cat, Adam Marcus, and Ivan Oransky. 2014. "The Peer-Review Scam." *Nature* 515(7528):480–482. doi:10.1038/515480a.

Greene, Sarah. 2016. "A New Metric: The Rapid Science Collaboration Score." *Retraction Watch*. February 5. http://retractionwatch.com/2016/02/05/a-new-metric-the-rapid-science-collaboration-score/.

Guengerich, F. Peter. 2015. "Some Changes in Submission and Handling Policies at the JBC." *Journal of Biological Chemistry* 290(46):28018–28019. doi:10.1074/jbc.E115.000001.

Keith, Ross. 2015. "Mirror Image in Plant Study Flagged on PubPeer Grows into Retraction." *Retraction Watch*. September 25. http://retractionwatch.com/2015/09/25/mirror-image-in-plant-study-flagged-on-pubpeer-grows-into-retraction/.

Lu, Susan Feng, Ginger Zhe Jin, Brian Uzzi, and Benjamin Jones. 2013. "The Retraction Penalty: Evidence from the Web of Science." *Scientific Reports* 3:3146. doi:10.1038/srep03146.

Marcus, Adam. 2014 "Pain Study Retracted for Bogus Data Is Second Withdrawal for University of Calgary Group." *Retraction Watch*. April 7. http://retractionwatch.com/2014/04/07/pain-study-retracted-for-bogus-data-is-second-withdrawal-for-university-of-calgary-group/.

Marcus, Adam, and Ivan Oransky. 2016. "Why Fake Data When You Can Fake a Scientist?" *Nautilus*. November 24. http://nautil.us/issue/42/fakes/why-fake-data-when-you-can-fake-a-scientist.

McCook, Alison. 2015. "The Consequences of Retraction: Do Scientists Forgive and Forget?" *Retraction Watch*. June 2016. http://retractionwatch.com/2015/06/16/life-after-retraction-in-many-cases-its-forgive-and-forget/.

McCook, Alison. 2016a. "Want to Correct the Scientific Literature? Good luck." *Retraction Watch*. February 3. http://retractionwatch.com/2016/02/03/six-barriers-to-correcting-the-scientific-record-qa-with-david-allison-and-andrew-brown/.

McCook, Alison. 2016b. "PubPeer Wins Appeal of Court Ruling to Unmask Commenters." *Retraction Watch*. December 7. http://retractionwatch.com/2016/12/07/pubpeer-wins-appeal-court-ruling-unmask-commenters/.

Munro, Margaret. 2014. "Prolific University of Calgary Doctor Resigned After Research Team Caught 'Manipulating' and Faking Data." *National Post*. September 8. http://news.nationalpost.com/news/canada/prolific-university-of-calgary-doctor-resigned-after-research-team-caught-manipulating-and-faking-data.

Oransky, Ivan. 2013. "Facing Legal Threats, Science Fraud Temporarily Suspends Posting." *Retraction Watch*. January 2. http://retractionwatch.com/2013/01/02/facing-legal-threats-science-fraud-temporarily-suspends-posting/.

Oransky, Ivan, and Adam Marcus. 2016. "Why Is It So Damn Hard to Get a Paper Retracted?" *STAT*. February 3. https://www.statnews.com/2016/02/03/paper-retracted/.

Retraction Watch. 2016a. "The Retraction Watch Leaderboard." Accessed May 15, 2016. http://retractionwatch.com/the-retraction-watch-leaderboard/.

Retraction Watch. 2016b. "Top 10 Most Highly Cited Retracted Papers." Accessed May 15, 2016. http://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/.

Steen, R. Grant, Arturo Casadevall, and Ferric C. Fang. 2013. "Why Has the Number of Scientific Retractions Increased?" *PLoS One* 8(7):e68397. doi:10.1371/journal.pone.0068397.

Van Noorden, Richard. 2011. "Science Publishing: The Trouble with Retractions." *Nature* 478:26–28. doi:10.1038/478026a.

Wakefield, Andrew J., Simon H. Murch, Andrew Anthony, John Linnell, D. M. Casson, Mohsin Malik, Mark Berelowitz, Amar P Dhillon, Michael A Thomson, Peter Harvey, Alan Valentine, Susan E Davies, and John A Walker-Smith. 1998. "Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, and Pervasive Developmental Disorder in Children." *The Lancet* 351(9103):637–641.