



Accountability in Research

Policies and Quality Assurance

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

What difference might retractions make? An estimate of the potential epistemic cost of retractions on meta-analyses

Daniele Fanelli, Julie Wong & David Moher

To cite this article: Daniele Fanelli, Julie Wong & David Moher (2021): What difference might retractions make? An estimate of the potential epistemic cost of retractions on meta-analyses, *Accountability in Research*, DOI: [10.1080/08989621.2021.1947810](https://doi.org/10.1080/08989621.2021.1947810)

To link to this article: <https://doi.org/10.1080/08989621.2021.1947810>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 14 Jul 2021.



[Submit your article to this journal](#)



Article views: 818



[View related articles](#)



[View Crossmark data](#)

What difference might retractions make? An estimate of the potential epistemic cost of retractions on meta-analyses

Daniele Fanelli ^a, Julie Wong^b, and David Moher^c

^aDepartment of Methodology, London School of Economics and Political Science, London, UK;

^bIndependent scholar; ^cCentre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada


ABSTRACT

The extent to which a retraction might require revising previous scientific estimates and beliefs – which we define as the epistemic cost – is unknown. We collected a sample of 229 meta-analyses published between 2013 and 2016 that had cited a retracted study, assessed whether this study was included in the meta-analytic estimate and, if so, re-calculated the summary effect size without it. The majority (68% of $N = 229$) of retractions had occurred at least one year prior to the publication of the citing meta-analysis. In 53% of these avoidable citations, the retracted study was cited as a candidate for inclusion, and only in 34% of these meta-analyses (13% of total) the study was explicitly excluded because it had been retracted. Meta-analyses that included retracted studies were published in journals with significantly lower impact factor. Summary estimates without the retracted study were lower than the original if the retraction was due to issues with data or results and higher otherwise, but the effect was small. We conclude that meta-analyses have a problematically high probability of citing retracted articles and of including them in their pooled summaries, but the overall epistemic cost is contained.

Retractions are a phenomenon of growing importance in science, but it is unclear if and to what extent they should be considered a hindrance to scientific knowledge and a symptom of falling standards, rather than a positive manifestation of scientific self-correction and integrity.

The number of retracted publications has grown from practically zero three decades ago to at least 1,970 items labeled as “retraction” in 2020 in the Web of Science database alone. This increase has generated concerns that misconduct itself might be rising in science. However, the number of retractions remains a comparatively small fraction (0.04%) of the literature (Brainard 2018), and multiple lines of evidence suggest that the growth in retractions results mainly or entirely from the expansion and strengthening

CONTACT Daniele Fanelli  email@danielefanelli.com  Department of Methodology, London School of Economics and Political Science, London, UK

 Supplemental data for this article can be accessed on the [publisher's website](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

of policies and practices to correct the literature (Fanelli 2013). The literature on COVID-19 offered a recent “natural experiment,” the results of which support this positive interpretation, by showing that retractions are occurring at faster and more efficient rates, and yet they remain relatively rare even when research is published under extreme pressures (in the Corona Central database, at the time of writing, COVID-related retractions are 97 out of 144,015 articles, or 0.067% (Coronacentral.ai 2021)). To the extent that growing retractions reflect improvements in scientific self-correction, then their increase should be celebrated as a positive development and further encouraged (Fanelli 2016; Fanelli, Ioannidis, and Goodman 2018).

Multiple independent studies have documented how the authors of a retracted study suffer a significant cost in terms of citations, productivity and funding (Lu et al. 2013,7; Mott, Fairhurst, and Torgerson 2018). However, these costs are only observed for authors of studies that are retracted due to misconduct. Authors of articles retracted for honest error appeared to suffer no negative consequence (Lu et al. 2013), and collaborators of authors of retracted articles suffer a loss of citations mainly when the retraction was due to misconduct (Mongeon and Larivière 2016, Hussinger and Pellens 2019; Azoulay, Bonatti, and Krieger 2017). If any reductions in citations, in productivity and in funding affect exclusively the individuals who committed scientific misconduct, then they do not quite represent a “cost” of retractions, but rather a fair sanction that the scientific community administers collectively to deserving individuals, creating deterrents that should be supported.

A clearer case may be made that retractions entail a waste of resources. However, the financial costs are surprisingly contained. A recent study, in particular, tried to estimate the financial impact of studies that were funded by the US National Institutes of Health and were retracted due to findings of misconduct. It concluded that the costs amount, in the least conservative estimates, to between 0.01% and 0.05% of the NIH budget, a figure that was deemed low in the authors’ own assessment (Stern et al. 2014). A prior case-study analysis conducted in 2010 concluded that, if the costs that were estimated for that case were extrapolated to all of the allegations made to the US Office of Research Integrity in their last reporting year, the costs of conducting all of those investigations would amount to 110 million US dollars (Michalek et al. 2010). This is clearly an implausible worse-case scenario, since most allegations do not require the same amount of investigation, if any at all, and yet the costs it suggests are only circa 0.35% of the NIH budget for 2010 (NIH 2020).

Financial costs of investigating misconduct aside, an argument is often made that other research resources are misdirected and wasted if they build upon retracted studies. Again, however, it is not obvious what the magnitude of this waste is, because it is proportional to how misleading and invalid the

retracted results are. Claims made by a retracted study could be scientifically correct and therefore not technically false or misleading even when the retraction is due data fabrication (Fanelli 2019).

Several studies have documented the fact that retracted articles continue to be cited after their retraction, although typically at much lower rates (Shuai et al. 2017; Mott, Fairhurst, and Torgerson 2018). Whereas part of these post-retraction citations could be negative or neutral (for example, they cite the retracted article as an example of misconduct), preliminary analyses suggest that most citations are actually positive, as if the retraction had not occurred (Bar-Ilan and Halevi, 2017; Budd, Coble, and Anderson 2011). If confirmed on a larger scale, these results could indicate a disturbing phenomenon, which may result in part from the indexing systems' inability to disseminate the retracted status of a reference (Schmidt 2018) and in part from willful actions of authors who ignore the retracted status of an article.

But to what extent is citing a retracted article damaging to scientific progress? It is unquestionably problematic from an ethical point of view, because it undermines the sanctioning role that a retraction ought to have on its authors. However, from a strictly scientific point of view the costs are less immediately obvious. The argument is similar to that made above for financial costs: citing a retracted article is harmful to scientific progress only to the extent that the retracted article reports incorrect or distorted results and that such results affect scientific beliefs.

We define as the “epistemic cost” of a retraction the extent to which the retraction distorts scientific knowledge. The magnitude of this cost has not yet been estimated.

Meta-analyses are an optimal tool to measure this impact, because they are designed to gauge the overall quantitative picture given by a literature. Therefore, meta-analyses represent key sources to inform scientific beliefs about the existence and magnitude of phenomena, and they are used extensively across the biological and social sciences (Fanelli, Costas, and Ioannidis 2017). If retracted studies bring no epistemic costs, removing them from a meta-analysis will have no impact on the meta-analytical results (i.e., estimate and precision). Vice-versa, a difference will be observed if and to the extent that retracted studies reported distorted information. The magnitude of this difference is an estimate of the epistemic cost of that retraction on that particular literature, because it represents the extent to which beliefs about a particular phenomenon need to be revised following the retraction.

A recent study examined the impact that a clinical trial that contained falsified data had on 22 meta-analyses that included it, and concluded that 10 (46%) of these meta-analyses had their results significantly changed (Garmendia et al. 2019). This result, however, cannot be generalized due to various limitations. It examined a single instance of data fabrication, even though retractions can occur for many different reasons. Further, it included

meta-analyses of the same or very similar research question. Furthermore, the fabrication of data was uncovered exclusively in a single Chinese clinical site and, once all Chinese data was excluded, the trial yielded statistically significant positive results (Seife 2015). Therefore, the average distorting effect that retractions exert on the scientific literature remains to be accurately assessed.

This study estimates the epistemic cost of a representative sample of recent retractions, by measuring the difference that these retractions make to the meta-analyses that included them. In a meta-assessment of bias in science, studies with a first author who had other articles retracted were significantly more likely to over-estimate effect sizes (Fanelli, Costas, and Ioannidis 2017). Therefore, our starting hypothesis was that meta-analyses that include retracted studies may overestimate effect sizes.

Materials and methods

Meta-analyses sampling strategy

The objective of our sampling strategy was to obtain a set of meta-analytical summaries (that is, quantitative weighted summaries of effect sizes extracted from multiple primary studies) that included a primary study that was recorded as retracted.

We first compiled a list of all records that were tagged as “retracted” in the Web of Science database (WOS) as of 16 December 2016. These records are reliably identified because their title in the WOS database is modified by the addition of the label “retracted article” or “retracted title.” The list was hand-inspected to exclude any false positives, leading to a total list of 3,834 records at the time (the WOS has recently updated its retraction tagging system and database, and now includes a larger number of retractions).

We subsequently compiled a list of all records in the WOS that cited any of the retracted records, obtaining an initial list of 83,946 records. We then restricted the above list to records that included “meta-analysis” OR “meta analysis” OR “systematic review” in the title, abstract or keywords, which resulted in 1,433 titles. For each of these records, the full list of cited references was retrieved, and the one or more references that had been retracted was identified by matching the document object identifier (DOI). Records for which the DOI did not identify any retracted cited reference were inspected by hand, to identify the WOS record of retracted cited reference.

The list was subsequently restricted to titles that included the Boolean string stated above exclusively in the abstract and that were published in 2015–2016. Following an earlier review of this article, the list was expanded to include the years 2013–2014.

Inclusion/exclusion of potentially relevant meta-analyses

Our objective to gauge the epistemic costs of retractions required a set of directly comparable studies that produced a weighted summary estimate of findings in primary studies. Therefore we adopted the following exclusion criteria:

- (1) Is limited to a systematic review, and not a formal meta-analysis
- (2) Is not a standard meta-analysis, in that it does not produce a single weighted pooled summary of two or more primary studies. This excludes network meta-analysis, Genome-Wide-Association-Studies, meta-analyses of neuroimaging data, microarray data, genomic data, etc.
- (3) Does not contain a usable summary of primary data in the full-text or accessible appendix. In particular, we required it to give a funnel plot or table containing data on each primary study's identity and reported effect size. An attempt was made to contact the authors of studies that met other inclusion criteria but lacked this information.
- (4) The retracted cited article is not one of the primary studies included in any of the weighted pool summaries presented (that is, it was cited for other reasons).

All but nine of the pdfs of the potentially relevant meta-analyses could be retrieved and manually inspected to determine exclusion based on these criteria.

Primary data extraction

From each included study, we identified the figure or table that reported the primary data for the meta-analysis. If the study contained more than one meta-analysis, we selected the one that cited the retracted study, and if more than one such meta-analyses was present we selected the first one shown in the publication.

For each of the selected meta-analyses, we recorded the pooled summary estimate and for each of the primary studies within each meta-analysis, we recorded the reported effect sizes and confidence intervals (or other measure of precision used, e.g., standard error, sample size). These numbers are provided in the publications and extracting them requires no subjective interpretation.

For each retracted primary study in our sample, we retrieved the text of the retraction note and we recorded the (one or more) reasons adduced in the note for the retraction. When in doubt about the reason, we compared the record with that in the Retraction Watch database (retractiondatabase.org).

org). Two studies that had been retracted for both data-related and non-data related reasons were categorized in the former category, in line with the hypothesis.

Calculating the impact of retractions

For each included meta-analysis, we calculated the summary effect size twice, i.e., with or without the retracted primary study. Where possible, meta-analytical summary estimates were obtained using the primary raw data in the form used by the original meta-analysis (e.g., number of events and totals, or means and standard deviation and sample size, etc.). Alternatively, summary estimates were simply re-calculated based on the processed numerical data (e.g., standardized mean difference \pm 95% confidence interval) reported in the figures (forest plots typically include data in numerical form) or table. Details of all calculations are given in the supplementary data file SI1.

To quantify the difference that removing the retracted study made to the meta-analysis, we adopted multiple strategies, because the included meta-analyses used different measures of effect size. In particular, we calculated the following measures:

- Effect size ratio: ratio, expressed in percentage, of the effect size (ES) with/without the retracted study. For each meta-analysis i , this was calculated as:

$$ESR_i = 100 \times \frac{ES(all\ primary\ studies)_i}{ES(retracted\ study\ removed)_i}$$

The ESR quantifies in very general terms how much larger or smaller the original summary effect size is relative to retraction-corrected one. It therefore quantifies in the most general sense the impact that the retracted study had on the meta-analysis, regardless of the metric used by the meta-analysis. An ESR of 100 indicates no difference, and an ESR greater than 100 indicates that including the retracted study led to a larger summary effect size in the meta-analysis.

- CI ratio: analogous to ESR, but using the size of the 95% Confidence Interval in place of ES.

To obtain a more metric-relevant estimate of the impact of a retraction, we limited our sample to meta-analyses with inter-convertible effect sizes (i.e., standardized mean difference, Hedges' g , both assumed to be equivalent to

a Cohen's d , and Odds Ratio, Peto Odds Ratio and Risk Ratio, all three assumed to be equivalent to Odds Ratio). On the resulting sub-set of meta-analyses we calculated the following metrics:

- ROR: ratio of odds ratios, which were inverted when necessary (IOR, see below), and calculated as:

$$ROR_i = \frac{IOR(all\ primary\ studies)_i}{IOR(retracted\ study\ removed)_i}$$

in which

$$IOR_i = \begin{cases} OR_i & \text{when } OR_i \geq 1 \\ OR_i^{-1} & \text{when } OR_i < 1 \end{cases}$$

The inversion is necessary to align the (eventual) biases in the same direction.

- DSMD: difference between standardized mean differences (SMD). Indicating the latter as d , this is calculated as:

$$DSMD_i = d(all\ primary\ studies)_i - d(retracted\ study\ removed)_i$$

- DCC: difference in correlation coefficient, r , calculated as:

$$DCC_i = r(all\ primary\ studies)_i - r(retracted\ study\ removed)_i$$

Data on meta-analytical studies

For each meta-analysis that cited a retracted study, we recorded the location (i.e., in the introduction, methods, results or discussion) and purpose of the citation – that is, whether the study was cited as a potentially relevant primary study for inclusion in the meta-analysis, or whether it was cited in the introduction, methods or discussion as background literature.

For each meta-analysis we also retrieved the full bibliometric record available in the Web of Science, on 12 March 2021.

Journal impact factors were obtained from the Web of Science's Journal Citation Reports for 2019 or for the last year for which they were available. Journals not included in the JCR were excluded from Impact factor analyses.

Journal discipline was attributed based on the Web of Science Essential Science Indicators system, which classifies journals in 22 disciplines.

Statistical analyses

All estimate and analyses in the text are obtained from generalized linear models, assuming a Gaussian distribution. To meet assumptions of normality, both the ESR metric and journal impact factor data were log transformed.

Results

Descriptive results

Of the 378 potentially relevant meta-analyses identified for the years 2013–2016 we could retrieve and inspect 369 full texts. Of these, 140 were not actually meta-analyses or were not simple weighted summaries, leaving a total of 229 relevant meta-analyses that cited somewhere in the text one or more retracted studies (Figure 1).

Why and how were retractions included?

Obviously, if the retraction occurred after, or even shortly before the publication of the meta-analysis, its inclusion would be explainable. However, in 156 meta-analyses in our sample (68%), the article had been retracted at least

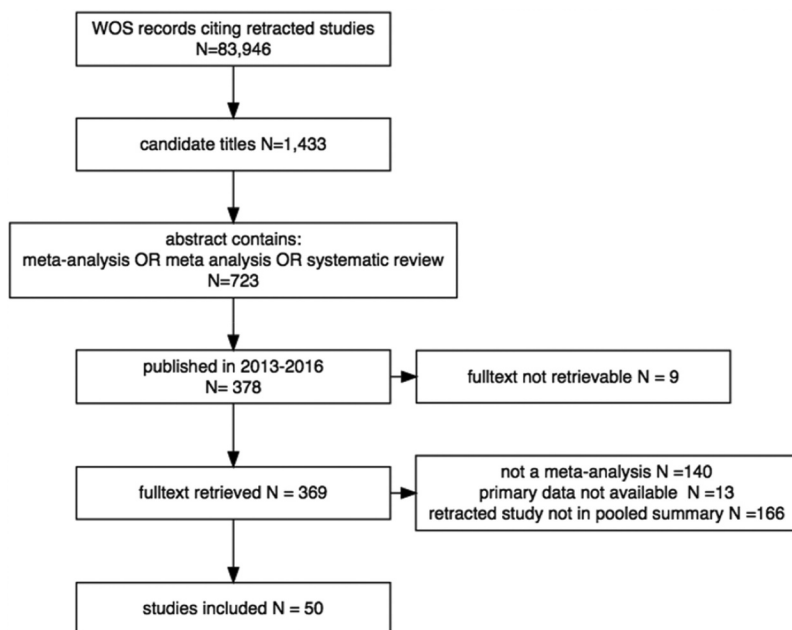


Figure 1. Flow chart of the sample retrieval process.

one year before the publication of the citing meta-analysis, suggesting that its citation could have been avoided. Indeed, on average, the cited retracted papers had been retracted circa 2.9 years before the publication of the citing meta-analyses.

Among these 156 meta-analyses with avoidable citations, 74 (47%) were citing the article in the introduction or discussion, as background information, and not because it was a candidate primary study for inclusion in the meta-analysis. Although we did not examine in depth the scope of these citations, most if not all of them appeared to be positive and failed to acknowledge the retracted status of the article.

Of the 82 remaining meta-analyses, 21 (26%, 13% of total) had excluded the study for reasons unrelated to the retraction, and 33 (40%, 21%) had included the retracted study in the meta-analysis. An additional 7 (9%, 4%) appeared, based on the full text, to have included the retracted study in the pooled summary, but we could not confirm this because these meta-analyses had not provided primary data and the corresponding authors did not respond to our request.

Therefore, only 21 meta-analyses had explicitly excluded the study because it had been retracted -or indeed appeared to openly acknowledge the retraction at all (13% of total avoidable citations, 34% of all meta-analyses that had selected a retracted study for inclusion).

The average journal impact factor of meta-analyses that excluded retracted studies was significantly higher than for meta-analyses that included them, controlling for discipline ($b = 0.533 \pm 0.235$, $P = 0.025$); meta-analyses that cited retracted studies for other reasons showed intermediate IF values and a broad confidence interval, suggesting that improper citations to retracted articles occur in journals of all disciplines and quality level (Figure 2).

What is the epistemic cost of including a retracted study?

In total, 50 meta-analyses were included in our main analysis, because their pooled summary contained a retracted study. Their main characteristics are reported in Table S1. Most of these meta-analyses were published in clinical medicine journals (35, 70%), and the rest were published in other areas of biology, with the exception of two meta-analyses in psychiatry/psychology and four in multidisciplinary journals (in particular, PLoS ONE).

These meta-analyses varied in size (i.e., number of primary studies included, range: 2–69), in number of citations received (range: 2–327) and in the metrics they used. The most common metric used in the sample was weighted mean difference ($n = 13$) and odds ratio ($n = 10$). All meta-analyses included a single retracted study except one, which included two retracted articles from the same author. The total number of retracted articles included by these meta-analyses was 40, indicating that ten meta-analyses had

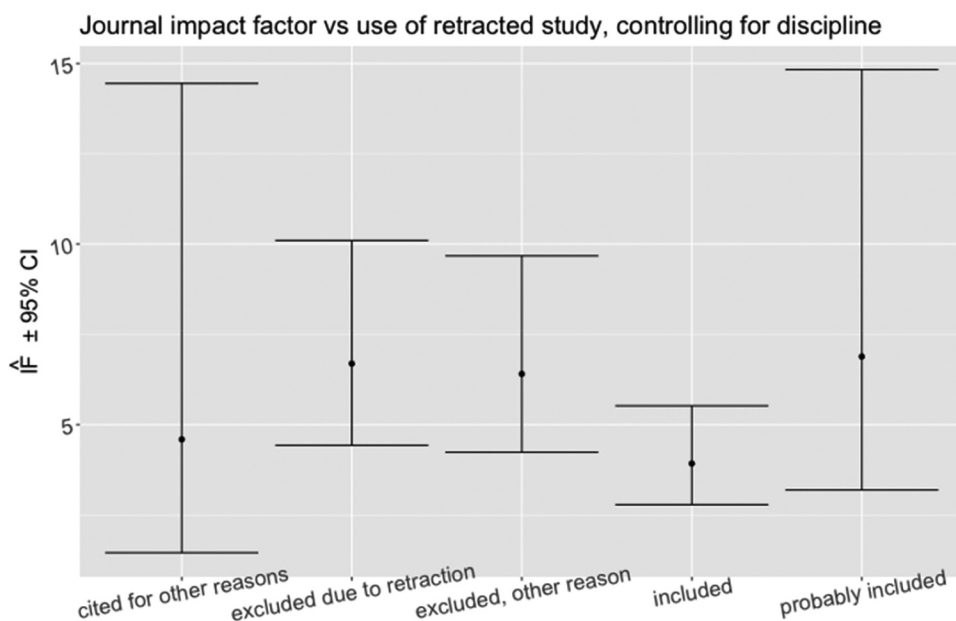


Figure 2. Estimated average journal impact factor of meta-analyses that cited a study that had been retracted at least one year prior to their publication, partitioned by why the study was cited and whether it was included in the pooled estimate. Values are mean and 95% CI, derived from a regression model that controlled for the discipline of the journal. Similar results are obtained not controlling for discipline. “Probably included” refers to meta-analyses that appeared to have included the retracted study in their pooled summary, but did not provide primary data that would allow verification.

included data from the same retracted study. There was no overlap between the authors of retracted studies and those of the meta-analyses in our sample.

The reasons for retraction, as described in the corresponding retraction notices, covered a varied spectrum of possible ethical infractions, including lack of ethical approval, authorship or peer-review issues, various forms of plagiarism, errors in data or methods, and data fabrication and falsification. Although in some cases more than one reason for retraction was indicated, retractions could be reliably separated into two binary categories: those that were (partially or entirely) due to issues with data, methods or results ($N = 25$) vs. those due to other causes ($N = 25$); and those that were connected to scientific misconduct ($N = 26$) vs. not ($N = 24$).

Recalculation of published effect sizes

In 49 of the 50 available meta-analyses, a pooled summary of the primary data set extracted had been calculated in the original publication. The recalculated effect sizes were generally in good agreement with values reported

in the publications: in 42 cases, the re-calculated values were identical or within $\pm 5\%$ of the published value, and the average percentage deviation was 0.57%. Discrepancies between original and re-calculated value were likely to be due to details of the calculations (e.g., rounding of numbers, particular statistical corrections) whose assessment was beyond the scope of this study. Because our objective was to assess the impact of retractions on pooled summaries, we used our re-calculated summaries as the baseline for comparison.

Effects of retraction on meta-analytical summary effect sizes

The ratio of effect sizes calculated with and without the retracted studies (henceforth, ESR) varied substantially, ranging between 48% and 270% circa. Across the entire sample, meta-analytical summaries with the retracted study included were not significantly larger than without it (mean ESR = 98.6, 95% CI = 91.7–106). ESR did not vary significantly across years of publication of the meta-analysis (log-likelihood ratio = 1.821, df = 3, $P = 0.6$) nor with size of meta-analysis ($b = 0.001 \pm 0.003$, $P = 0.839$). The effect size calculated without the retracted study was within the confidence interval calculated with all studies included in 48 of the 50 cases (96%).

However ESR varied significantly depending on the nature of the retraction (Figure 3). The ESR was significantly larger for meta-analyses that included a study that was retracted due to issues with data, methods or results relative to the others (GLM controlling for meta-analysis size: $b = 0.16 \pm 0.071$, $P = 0.027$; mean and 95%CI, respectively: 106.8 [92.5–123.4] vs. 90.9 [82.1–100.7]). The ESR was not significantly different between meta-analyses in which the included study had been retracted due to misconduct relative to other reasons (controlling for MA size and data-relevance of retraction, $b = 0.052 \pm 0.072$, $P = 0.48$). The difference between data-related and non-data-related retractions was largest when multiple confounding factors were controlled for, in a model adjusting for misconduct status, meta-analysis size, discipline and year of meta-analysis ($b = 0.212 \pm 0.085$, $P = 0.018$).

To gain a more accurate estimate of the magnitude of this effect, we limited the analysis to meta-analyses with inter-transformable metrics and we ran a multiple linear regression. Controlling for meta-analysis size, discipline and whether the retraction was due to misconduct, estimates suggest a difference of about 28% in ROR, 14% in DSMD and 6.2% in DCC.

Confidence intervals tended to expand when the retracted study was removed, as would be expected, but the effect was very modest and not statistically significant (estimated mean CI ratio and 95%: 98.3 [85.56; 112.93]). The CI ratio did not differ significantly according to meta-

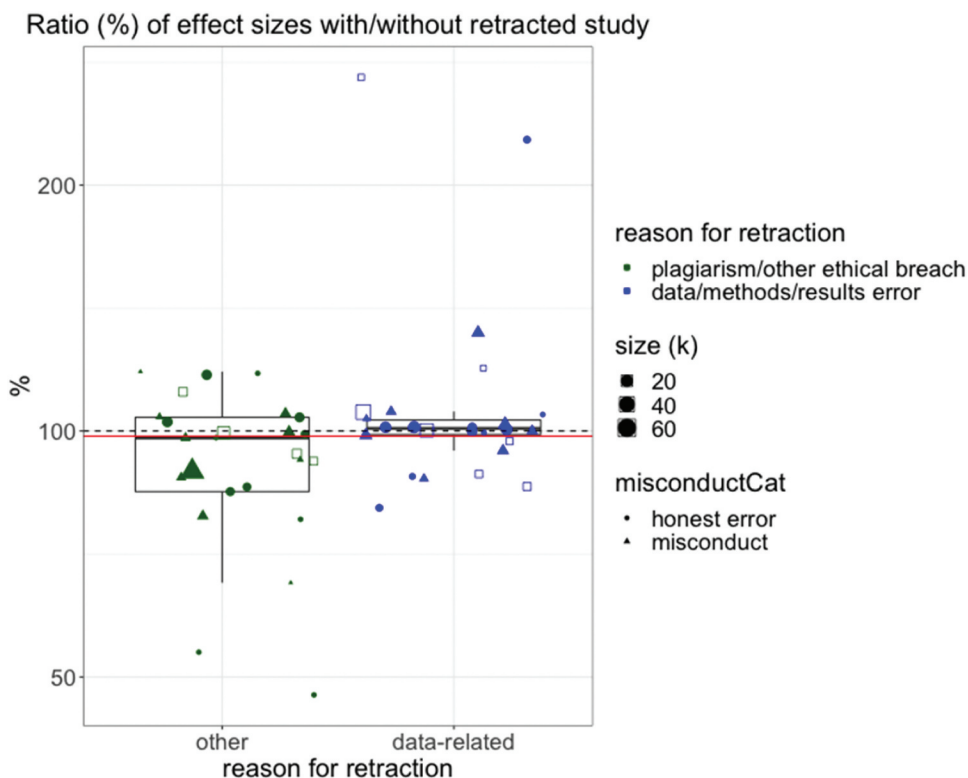


Figure 3. Relative size, in percentage, of meta-analytical pooled effect sizes calculated including and excluding the retracted study. Values of 100% (dashed line) indicate identical effect sizes, whereas values above 100% indicate that the meta-analysis with the retracted study yielded a larger effect size than without it. Empty squares indicate meta-analyses that included the same retracted study as another meta-analysis. Red line indicates the overall estimate.

analysis size, or data-relevance of retraction or misconduct status (respectively: $b = 0.007 \pm 0.006$ $P = 0.262$, $b = -0.177 \pm 0.138$ $P = 0.208$, $b = 0.026 \pm 0.139$ $P = 0.851$).

The ESR was negatively (and marginally significantly) associated with impact factor of the journal in which the meta-analysis was published, both alone (GLM with log-transformed IF: $b = -0.071 \pm 0.040$, $P = 0.077$) and after controlling for meta-analysis size, data-related retraction, misconduct, and discipline ($b = -0.080 \pm 0.042$, $P = 0.066$). Based on the latter model, we can estimate that for every 1-point reduction in impact factor the ESR increased on average by 8.2% [95%: -0.003%; 17.6%].

Confounding factors and robustness tests

As an alternative, more robust and conservative analysis, we ran a multiple logistic regression model on the log odds of ESR being larger than 100%. This analysis yielded similar results to our central finding, suggesting that

controlling for meta-analysis size, discipline, meta-analysis publication year, and whether the retraction was due to misconduct, the odds of having ESR >100% were about four times higher for retractions due to data, methods or results compared to the others (logistic regression: $b = 1.431 \pm 0.682$, $P = 0.0357$, odds ratio = 4.18).

All results described above were obtained including all 50 meta-analyses, regardless if they shared a retraction. Substantially equivalent results were obtained if only one of any two meta-analyses sharing a retraction was included (see [Figure 1](#)).

Our main response variable had inverted all ratio measures for which the pooled estimate was below 1 (see methods). Omitting this transformation in the analysis yielded similar estimates for the overall ESR (99.%) but virtually nullified the difference in ESR between retractions due to data issues vs. not (e.g., controlling for all other confounding factors, $b = 0.058 \pm 0.082$, $P = 0.484$). Interestingly, this uninverted ESR tended to be higher for misconduct retractions ($b = 0.129 \pm 0.073$, $P = 0.081$), less so when controlling for the other confounders ($b = 0.133 \pm 0.079$, $P = 0.101$).

Our main result was only modestly robust to outliers: if we removed from the analysis two data points whose magnitude of the log-transformed ESR was larger than 3 standard deviations, results-related retractions still had higher average ESR, but the effect was less than half in magnitude and no longer rejecting the null ($b = 0.088 \pm 0.051$, $P = 0.09$), especially once controlling for all the confounding factors discussed above ($b = 0.080 \pm 0.065$, $P = 0.227$).

Discussion

We identified a sample of meta-analyses published between 2013 and 2016 that had cited one or more retracted articles, and assessed if theses had been included in the pooled summaries and, if so, with what results.

We found that the majority of retractions had occurred before the publication of the citing meta-analyses, and yet only about 13% of the latter had explicitly excluded the article because it was retracted. Meta-analyses published in low-ranking journals were more likely to include previously retracted studies. To any extent that Impact Factor metrics reflect the quality of a journal, this suggests that the risk for a meta-analysis to include a retracted study is higher for low-quality journals. However, general citations to retracted studies occurred in journals of all levels ([Figure 2](#)), showing such mis-citations to be ubiquitous.

The epistemic cost of retractions on meta-analytical pooled summaries was on the average modest, although likely to be variable and context-dependent. On the whole, pooled estimates calculated without the retracted study were similar in magnitude and direction to the original, and in 96% of cases were

within the original 95% confidence interval. Therefore, these retractions would not have drastically altered the conclusions originally drawn from the meta-analyses, at least to the extent that such conclusions had taken confidence intervals into account – as they should.

Taken as an average across the sample, results failed to support the hypothesis that retracted studies would generally distort meta-analyses in the direction of an over-estimation of effect size. However, we found that the impact of retractions varied with the cause of retraction. Studies retracted due to problems with the data, methods, or results were associated with higher original pooled effect size, whereas studies retracted for plagiarism and other non-data-related issues had lowered the original estimate. The most accurate and conservative estimates in our samples suggest an average difference between these two groups of between 6% and 28%, depending the metric considered.

Before discussing the implications of these results, a few important limitations need highlighting. The first limitation is the relatively small size of our sample ($N = 50$). Combined with the methodological heterogeneity of meta-analyses (which differed in various characteristics including the number of primary studies included and the statistical metrics used), and with the strong influence exerted by extreme values, a small sample size makes our results a preliminary estimate of the *potential* impact of retractions. Studies that intend to measure accurately the extent to which retractions are actually distorting the evidence of specific fields should repeat our methods in larger and more homogeneous samples of meta-analyses taken from specific fields. A larger and more homogeneous sample might also reveal other plausible patterns that were not observed in this study, for example a difference between retractions due to misconduct vs. honest error.

A second possible limitation in our analysis is the reliance, when classifying the causes of a retraction, on retraction notices. These are known to be an imperfect tool because they may depict the causes of a retraction inaccurately, in particular by under-reporting the occurrence of scientific misconduct (Fang, Steen, and Casadevall 2012). This limitation might weaken our estimate of the difference between retractions due to misconduct vs. the rest. However, it is unlikely to affect our main observation of a difference between issues relating to data versus not, because a retraction notice is very unlikely to misrepresent this aspect (for example, it is unlikely to falsely report as data fabrication what was in reality a case of plagiarism).

A third possible limitation is that our sample might not have captured citations to all articles that have been retracted, especially if these were not marked in an identifiable way in the WOS. However, we expect such errors to be relatively rare, and therefore unlikely to affect our results. Moreover, any classification errors would be unrelated to the type of retraction and to

whether or not an article was included in a meta-analysis. Therefore, this limitation is likely to merely add noise to our analysis, reducing statistical power and making, once more, our results more conservative.

Finally, we should remark that most secondary patterns observed (e.g., differences in impact factor and between types of retractions) had not been explicitly hypothesized at the start of the study, and thus should be considered exploratory results. However, these patterns make good theoretical sense, in retrospect, which reduces the risk that the result is spurious (Rubin 2017).

Two main conclusions can be derived from these results, sending both a re-assuring and a concerning message, respectively. The re-assuring message is that retractions as a whole appear to have a modest epistemic cost on meta-analyses. Meta-analytical methods are protected from the epistemic cost of retractions for at least two reasons. Firstly, because the very logic of pooling different studies makes results generally robust to any single study being flawed. Secondly, because properly specified inclusion criteria are often able to exclude flawed studies regardless of their official retraction status. Many of the retracted studies in our sample, for example, had been excluded from meta-analyses due to “data overlap” or “lack of sufficient information” – in other words, for issues that foreshadowed their flaws, regardless of whether they had been retracted or not.

The troubling message, however, is that a large number of retracted studies are cited and included in meta-analyses, supporting previous concerns that scientific self-correction is inefficient in this regard (Bar-Ilan and Halevi, 2017; Budd, Coble, and Anderson 2011). Meta-analyses are intended to be very rigorous analyses of the published literature, and it could have been hypothesized that they would tend to avoid the erroneous citation patterns observed in ordinary research articles. Our results suggest otherwise. Even though our results suggest a modest epistemic cost on average, the continuing positive use of retracted studies defies the purpose of retractions. Furthermore, our results suggest that the magnitude and direction of the epistemic cost may vary depending on the type of retraction and other context-specific factors (e.g., discipline, quality of meta-analysis). In fields where extreme precision is necessary, the impact of a retraction could be significant, and hard to predict.

Since most retractions in our sample had occurred long before the publication of the citing meta-analyses, this problem is largely avoidable, and our results support calls for improving methods of indexing and signaling retractions to reduce their epistemic cost (Schmidt 2018). Better indexing methods ought to be adopted not just by standardized bibliographic databases but also by Google Scholar and other less structured public literature sources, which may be increasingly used by researchers and thus contribute to expanding the problem of mis-citations.

The finding that studies retracted due to data, methods or results might over-estimate effect sizes, supports the hypothesis that many forms of error, bias and misconduct in science are directed at producing “positive” results, and therefore generally lead to a literature that over-estimates the significance and magnitude of effects. This hypothesis was also supported by a previous meta-meta-analytical study, which found that primary studies whose first authors had articles retracted reported significantly more extreme effect sizes (Fanelli, Costas, and Ioannidis 2017). This hypothesis was also strongly suggested by intuition and experience, leading several authors to assume that bias in research is primarily a bias toward false-positives (e.g., Ioannidis 2005). Therefore, our results corroborate previous evidence and intuitive assumptions about a link between (certain kinds of) problematic research and the rate of false positive results in the literature, and offer a preliminary estimate of the magnitude of this effect.

The finding that studies retracted due to non-data related issues have the opposite effect to the others support the notion that retractions are not all the same, and that policies and practices may benefit from operating distinctions between them. Providing clearer and accessible information on the nature of a retraction would not only set fairer incentives for researchers to self-correct their own mistakes (Fanelli 2016; Fanelli, Ioannidis, and Goodman 2018), but it would also help to reevaluate a scientific literature after errors or misconduct come to light. Scientists’ beliefs about the magnitude of a phenomenon might need to be corrected downwards if a retraction is due to problems with data, and upwards otherwise.

In conclusion, our findings support concerns that the retracted status of articles is too often overlooked by researchers, and yet suggest that the epistemic cost of retractions on meta-analyses is relatively contained and very context-dependent. The cumulative nature of knowledge produced by meta-analyses makes them relatively robust to the impact of retractions. However, in some fields even small changes in pooled estimates may have profound implications, generating problems that a more efficient retraction system could largely prevent.

Acknowledgments

Author contributions: DF: Conceptualization, Data Curation, Data Collection, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation; DM: Conceptualization, Validation, Writing – Review & Editing; JW: Data Collection, Writing – Review & Editing.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Daniele Fanelli  <http://orcid.org/0000-0003-1780-1958>

References

- Azoulay, P., A. Bonatti, and J. L. Krieger. 2017. "The Career Effects of Scandal: Evidence from Scientific Retractions." *Research Policy* 46 (9): 1552–1569. doi:10.1016/j.respol.2018.01.012.
- Bar-Ilan, J., and G. Halevi. 2017. "Post Retraction Citations in Context: A Case Study." *Scientometrics* 113 (1): 547–565. doi:10.1007/s11192-017-2242-0.
- Brainard, J. 2018. "What a Massive Database of Retracted Papers Reveals about Science Publishing's 'Death Penalty'." *Science* (80-). doi:10.1126/science.aav8384.
- Budd, J. M., Z. C. Coble, and K. M. Anderson. 2011. "Retracted Publications in Biomedicine: Cause for Concern." Declaration of Interdependence: The Proceedings of the ACRL 2011 Conference, pp. 390–395. Philadelphia, PA, March 30–April 2. http://0-www.ala.org.sapl.sat.lib.tx.us/ala/mgrps/divs/acrl/events/national/2011/papers/retracted_publicatio.pdf
- Coronacentral.ai [Internet]. Accessed 14 April 2021. <https://coronacentral.ai/>
- Fanelli, D. 2013. "Why Growing Retractions Are (Mostly) a Good Sign". *PLoS Med.* 10: 1–6. doi:10.1371/journal.pmed.1001563
- Fanelli, D. 2016. "Set up a 'Self-retraction' System for Honest Errors." *Nature* 531 (7595): 415. doi:10.1038/531415a.
- Fanelli, D. 2019. "A Theory and Methodology to Quantify Knowledge." *Royal Society Open Science* 6 (4): 181055. doi:10.1098/rsos.181055.
- Fanelli, D., J. P. A. Ioannidis, and S. Goodman. 2018. "Improving the Integrity of Published Science: An Expanded Taxonomy of Retractions and Corrections." *European Journal of Clinical Investigation* 48 (4): e12898. doi:10.1111/eci.12898.
- Fanelli, D., R. Costas, and J. P. A. Ioannidis. 2017. "Meta-assessment of Bias in Science." *Proceedings of the National Academy of Sciences* 114 (14): 3714–3719. doi:10.1007/s11192-017-2242-0.
- Fang, F. C., R. G. Steen, and A. Casadevall. 2012. "Misconduct Accounts for the Majority of Retracted Scientific Publications." *Proceedings of the National Academy of Sciences of the United States of America* 109 (42): 17028–17033. doi:10.1073/pnas.1212247109.
- Garmendia, C. A., L. Nassar Gorra, A. L. Rodriguez, M. J. Trepka, E. Veledar, and P. Madhivanan. 2019. "Evaluation of the Inclusion of Studies Identified by the FDA as Having Falsified Data in the Results of Meta-analyses: The Example of the Apixaban Trials." *JAMA Internal Medicine* 179 (4): 582–584. doi:10.1001/jamainternmed.2018.7661.
- Hussinger, K., and M. Pellens. 2019. "Guilt by Association: How Scientific Misconduct Harms Prior Collaborators." *Research Policy* 48 (2): 516–530. doi:10.1016/j.respol.2018.01.012.
- Ioannidis, J. P. A. 2005. "Why Most Published Research Findings are False." *PLoS Med. Public Library of Science* 2: e124.
- Lu, S. F., G. Z. Jin, B. Uzzi, and B. Jones. 2013. "The Retraction Penalty: Evidence from the Web of Science." *Scientific Reports* 3 (1): 5. doi:10.1073/pnas.1212247109.
- Michalek, A. M., A. D. Hutson, C. P. Wicher, and D. L. Trump. 2010. "The Costs and Underappreciated Consequences of Research Misconduct: A Case Study." *PLoS Medicine* 7. doi:10.1371/journal.pmed.1000318.
- Mongeon, P., and V. Larivière. 2016. "Costly Collaborations: The Impact of Scientific Fraud on Co-authors' Careers." *Journal of the Association for Information Science and Technology* 67 (3): 535–542. doi:10.1002/asi.23421.

- Mott, A., C. Fairhurst, and D. Torgerson. 2018. "Assessing the Impact of Retraction on the Citation of Randomized Controlled Trial Reports: An Interrupted Time-series Analysis." *Journal of Health Services Research & Policy* 24 (1): 44–51. doi:[10.1177/1355819618797965](https://doi.org/10.1177/1355819618797965).
- NIH. "The NIH Almanac [Internet]." Accessed 27 March 2020. <https://www.nih.gov/about-nih/what-we-do/nih-almanac/appropriations-section-2>
- Rubin, M. 2017. "When Does HARKing Hurt? Identifying When Different Types of Undisclosed Post Hoc Hypothesizing Harm Scientific Progress." *Review of General Psychology : Journal of Division 1, of the American Psychological Association* 21 (4): 308–320. doi:[10.1037/gpr0000128](https://doi.org/10.1037/gpr0000128).
- Schmidt, M. 2018. "An Analysis of the Validity of Retraction Annotation in Pubmed and the Web of Science." *Journal of the Association for Information Science and Technology* 69 (2): 318–328. doi:[10.1002/asi.23913](https://doi.org/10.1002/asi.23913).
- Seife, C. Research Misconduct Identified by the US Food and Drug Administration. *JAMA Intern Med.* 2015;175: 567. doi:[10.1001/jamainternmed.2014.7774](https://doi.org/10.1001/jamainternmed.2014.7774)
- Shuai, X., J. Rollins, I. Moulinier, T. Custis, M. Edmunds, and F. Schilder. 2017. "A Multidimensional Investigation of the Effects of Publication Retraction on Scholarly Impact." *Journal of the Association for Information Science and Technology* 68 (9): 2225–2236. doi:[10.1002/asi.23826](https://doi.org/10.1002/asi.23826).
- Stern, A. M., A. Casadevall, R. G. Steen, and F. C. Fang. 2014. "Financial Costs and Personal Consequences of Research Misconduct Resulting in Retracted Publications." In *Elife*, edited by P. Rodgers, Vol. 3, e02956. eLife Sciences Publications. doi:[10.7554/eLife.02956](https://doi.org/10.7554/eLife.02956).