Retractions in the research literature: misconduct or mistakes?

Sara B Nath, Steven C Marcus and Benjamin G Druss

Reasons for retractions in the research literature, according to the United States National Library of Medicine, include "pervasive error or unsubstantiated or irreproducible data" in an article.[1] Retractions typically indicate a problem with a study that is of sufficient magnitude to completely invalidate its findings. As such, they represent a threat both to the integrity of the scientific literature and to any future studies based on the erroneous conclusions. Most of the previous work on retractions in the biomedical literature has focused on the problem of continuing citation, as it is common for articles to be cited long after they have been formally retracted.[2-4]

In clinical medicine, in the relatively young field of patient safety, it has been found that medical errors typically do not result from physician malfeasance, incompetence, or ignorance, but rather are caused by unintentional mistakes compounded by system failures.[5]

Editorial policies, on the other hand, have generally focused more on scientific misconduct than unintentional mistakes. For instance, the International Committee of Medical Journal Editors, in its statement on retractions, says, "It is conceivable that an error could be so serious as to vitiate the entire body of the work, but this is unlikely and should be handled by editors and authors on an individual basis".[6] Several studies have focused on understanding the causes and consequences of egregious cases of scientific misconduct.[2,7,8] However, far less is known about retraction due to unintentional errors.

We examined the problem of error in scientific literature by focusing on understanding (i) how commonly articles are retracted on the basis of unintentional mistakes, and (ii) how these articles differ from those retracted for scientific misconduct in authorship and timing of the retraction statements, and the type of journals publishing retractions.

METHODS

Using the MEDLINE database, we identified all English language clinical or basic science articles that had been retracted between 1982 (the earliest date for which there was a formal retraction policy at the National Library of Medicine) and 2002 (the latest date for which there was sufficient time for published arti-

ABSTRACT

Objective: To determine how commonly articles are retracted on the basis of unintentional mistakes, and whether these articles differ from those retracted for scientific misconduct in authorship, funding, type of study, publication, and time to retraction.

Data source and study selection: All retractions of English language publications indexed in MEDLINE between 1982 and 2002 were extracted.

Data extraction: Two reviewers categorised the reasons for retraction of each article as misconduct (falsification, fabrication, or plagiarism) or unintentional error (mistakes in sampling, procedures, or data analysis; failure to reproduce findings; accidental omission of information about methods or data analysis).

Data synthesis: Of the 395 articles retracted between 1982 and 2002, 107 (27.1%) were retracted because of scientific misconduct, 244 (61.8%) because of unintentional errors, and 44 (11.1%) could not be categorised. Compared with articles retracted because of misconduct, articles with unintentional mistakes were more likely to have multiple authors, no reported funding source, and to be published in frequently cited journals. They were more likely to be retracted by the author(s) of the article, and the retraction was more likely to occur more promptly (mean, 2.0 years; 95% CI, 1.8–2.2) than articles withdrawn because of misconduct (mean, 3.3 years; 95% CI, 2.7–3.9) ($P < 0.05$ for all comparisons).

Conclusions: Retractions in the biomedical literature were more than twice as likely to result from unintentional mistakes than from scientific misconduct. The different characteristics of articles retracted for misconduct and for mistakes reflect distinct causes and, potentially, distinct solutions.

MJA 2006; 185: 152–154

cles to be retracted). All examples of publication type "retracted publication" were identified, as were the retracting letters or editorials (identified as "retraction of publication"). If a statement retracted more than one article, the retracted articles were used as the unit of analysis rather than the retraction statement.

The full text of the retractions and the retracted articles were reviewed separately by two of us (S B N, B G D).

Classification of retractions

The articles were grouped into one of three categories.

• Misconduct was classified, using the definitions of scientific misconduct from the US Office of Science and Technology Policy, as either fabrication (making up data or results and recording or reporting on them); falsification (manipulating research materials, equipment, or processes; or changing or omitting data or results such that the research is not accurately represented in the research record); or plagiarism (the appropriation of another person's ideas, processes, results, or words without giving appropriate credit).[9]

• Mistakes were defined as reported errors in sampling, procedures, or data analysis; failure to reproduce findings or accidental omission of key information from methods or analysis.

• Retractions impossible to classify as either misconduct or mistakes made up a third category. These retractions did not cite a reason for the action, or the information provided was insufficient to distinguish whether the retraction was prompted by misconduct or mistakes.

Characteristics of retracted articles

The characteristics of the retracted articles considered in the analyses included type of journal in which they were published, number of authors, funding source (any or no reported funding source), type of study (clinical or basic science studies), and date of publication (before or after 1991, the midpoint of the study period).

To characterise the journals in which retracted articles were published, we used the Institute for Scientific Information (ISI) impact factor, which reflects the number of citations to articles in a journal divided by the

total number of articles published in that journal during a specific time period.[10] A dichotomous variable was created to identify whether the articles were published in one of the top 100 ISI journals.

Two additional variables were abstracted based on the notice of retraction (as opposed to the retracted article); namely, whether one or more of the authors of the retracted article also wrote the retraction, and the period between the initial article and the retraction.

Statistical analysis

Bivariate analyses (χ^2, *t* tests, ANOVA) were used to compare each of the above characteristics between the subset of articles that could be classified as either misconduct or mistakes ($n = 351$). We used SAS, version 9.1 (SAS Institute Inc, Cary, NC, USA) for all analyses.

RESULTS

The standard κ coefficient for the agreement between the two abstracters was 0.875 out of a possible 1, indicating a high level of inter-rater reliability.[11]

Number and types of retractions

Between 1982 and 2002, 395 articles indexed in MEDLINE were retracted. Of these, 107 (27.1%) were classified as scientific misconduct. An example of this was a case in which a formal university investigation found that a survey interviewer had fabricated interview records in a study of second-hand smoke and asthma.[12]

A much larger proportion of retractions, (244; 61.8%) fell into the category of mistakes. One example of a mistake was an article published in 2002 in *Science* reporting neurotoxicity related to ecstasy use in primates. A detailed retraction published the following year reported that the bottle containing the sample had been mislabelled and in fact contained not ecstasy, but methamphetamine.[13]

For 44 retractions (11.1%), there was insufficient information to categorise the retraction as either misconduct or mistake. The most common reason for inability to classify was that there was no information at all about the reason for the retraction.

Comparison of retracted articles: misconduct versus mistakes ($n = 351$)

The results are summarised in the Box. Compared with articles retracted for reasons of misconduct, those retracted for mistakes were less likely to be written by a single author (5.7% v 10.5%; $\chi^2_2 = 6.2$; $P = 0.04$) and more likely to have more than five authors

Comparison of articles retracted due to scientific misconduct and unintentional errors ($n = 351$)

Characteristic of retracted article	Misconduct ($n = 107$)	Mistakes ($n = 244$)	P
Number of authors			0.04
1	10.5%	5.7%	
2–4	61.1%	53.2%	
5 or more	28.6%	41.1%	
No reported funding source	40.5%	59.4%	0.005
Clinical study (human subjects)	69.3%	65.6%	0.27
Published after 1991	58.9%	65.6%	0.23
Journal ranking by impact factor — in top 100 (as ranked by ISI)	34.6%	47.5%	0.02
Retraction written by author of the initial manuscript	35.2%	90.2%	< 0.001
Time between initial article and retraction (mean, years)	3.3 (95% CI, 2.7–3.9)	2.0 (95% CI, 1.8–2.2)	< 0.001

ISI = Institute for Scientific Information. ◆

(41.1% v 28.6%; $\chi^2_2 = 6.2$; $P = 0.04$). Mistakes were more likely to be in articles with no reported funding source (59.4% v 40.5%; odds ratio [OR], 2.40; 95% CI, 1.40–3.80; $\chi^2_1 = 7.74$; $P = 0.005$). There were no differences in types of retractions based on the date of publication (65.6% v 58.9%; $\chi^2_1 = 1.4$; $P = 0.23$) or in clinical compared with basic science studies (65.6% v 69.3%; $\chi^2_1 = 1.2$; $P = 0.27$).

As anticipated, mistakes were substantially more likely than misconduct to be reported by an author of the initial manuscript (90.2% v 35.2%; OR, 16.6; 95% CI, 9.0–31.0; $\chi^2_1 = 114.1$; $P < 0.001$). The mean time lapse between the original article and the retraction was more than a year shorter for mistakes (2.0 years; 95% CI, 1.8–2.2) than for misconduct (3.3 years; 95% CI, 2.7–3.9; $t = 4.67$, df = 349; $P < 0.001$).

DISCUSSION

In our analysis, unintentional mistakes were more commonly given as a reason for article retractions than scientific misconduct. However, retractions, as a whole, are quite rare. The 395 retractions identified represented a tiny fraction of the nine million articles in the MEDLINE database between 1982 and 2002.

There is some evidence that these retractions, particularly those due to mistakes, represent only the "tip of the iceberg". One piece of evidence is that both types of retracted articles were far more likely to be published in highly visible and frequently cited journals. In the broader literature, 4.2% of articles are published in the top 100 ISI journals.[14] Publication in one of these high-impact journals was eight times more common for articles retracted due to misconduct, and more than 11 times more likely for articles retracted due to mistakes.

The three journals with the highest number of retractions in this study were *Science*, *Proceedings of the National Academy of Sciences*, and *Nature*. It seems highly unlikely that these journals are prone to publishing shoddy research. Instead, this elevated error rate may reflect the high level of post-publication scrutiny received by the articles in these journals. It is likely to be easier for errors to slip by undetected in less widely read and cited journals. In addition, the complexity and rigour associated with studies published in these journals may lead to a higher risk for error in implementing and replicating the research. Furthermore, the large volume of articles published in these journals may naturally increase the rate of error among them.

A second piece of evidence to suggest that errors may be more common than the number of retractions indicates comes from examining the rate of errata — mistakes that warrant correction but are not of sufficient magnitude to require a full retraction.[1] Although some errata are related to only minor flaws such as typographical errors, during the study period there were 2772 errata published, more than seven times the number of retractions published in that time. Thus, only a fraction of research mistakes prove to be damaging enough to the integrity of a study to require a full-scale retraction.

However, these are only limited clues about the broader problem of inaccuracies in published scientific literature. Ascertaining

the true prevalence of these errors would require an approach similar to that used in epidemiological studies of illness. This approach would include analysis of random samples of articles from the scientific literature. Smaller-scale efforts to reanalyse study data have found that it is quite common to obtain results that are strikingly different from those in the original reports.[15]

The different characteristics of articles retracted for misconduct and for mistakes found in this study suggest that these represent distinct problems, with distinct causes, and, potentially, distinct solutions. Whereas single authorship, with its lack of accountability and oversight by colleagues, may raise the risk of misconduct, multiple authorship may increase the potential both for mistakes to occur and the opportunity for co-authors to detect them after publication. Similarly, having a source of funding may help ensure the rigour of the study to prevent mistakes, but this may also be a mechanism by which instances of misconduct are detected. Unfortunately, successful application for research support is unlikely to prevent or mitigate mistakes that occur several years later. Finally, cases of misconduct, which often require lengthy investigations, take longer to result in retraction statements than do mistakes.

There are lessons from the field of patient safety that can be applied to the problem of research errors. The patient safety field was built on the broader empirical and conceptual literature on human error.[5,16] Similarly, lessons from that literature can be applied to understanding research error.[17] At least three lessons from that literature may be particularly relevant to detecting and mitigating errors in biomedical literature.

The first lesson is the simple fact that humans, even diligent, meticulous and highly trained professionals, make mistakes. One of the main contributions of research in patient safety has simply been documenting the fact that these mistakes are more common than our experience leads us to believe. Approaching this problem for the research literature will require, first and foremost, fostering an environment of transparency in which authors and journals feel comfortable in reporting errors when they occur. A certain percentage of retractions in this study did not list reasons for the retraction. Journal editors may be reluctant to print retractions with sufficient information because of fears of litigation from authors.[18] This shows some discomfort on the part of authors and journals in admitting mistakes. However, the impact of published retractions is in part determined by researchers seeking them out. In 1987, Garfield commented that scientists should make a habit of searching for errata and retractions when performing literature searches.[19] Today, many search engines make this task much easier by allowing for easy access to retractions and corrections.

The second lesson is that different types of errors require different strategies for detection and mitigation of their consequences. For instance, Reason describes two distinct categories of errors: errors in planning, and errors in execution.[16] The peer review process is mainly designed to assess the former, assuring, for instance, that study designs are appropriate and that correct statistical tests are used. Many journals use statisticians in the peer review process to achieve this aim. It is more difficult to ascertain the problem of execution; for example, whether there were problems in data collection or analysis. For example, members of a peer review panel don't have access to raw data to determine if the results of a study are flawed. As a result, detecting problems in execution requires either re-analyses of key data, or attempts by other authors to replicate findings once they are published.

The final, and most important, lesson to be learned from the human error literature is that strategies for reducing error are very different from those used to detect and handle scientific misconduct. Whereas "naming, shaming and blaming" may be appropriate for dealing with scientific misconduct, these approaches are not effective, and may even be counterproductive, in reducing unintentional errors. Reducing errors requires a commitment to building systems that can prevent, detect, and mitigate the effects of errors when they occur.[5] Ultimately, research mistakes, like all human errors, must be seen not as sources of embarrassment or failure, but rather as opportunities for learning and improvement.

COMPETING INTERESTS

None identified.

AUTHOR DETAILS

Sara B Nath, PhD, MSW, Assistant Professor[1]
Steven C Marcus, PhD, Research Associate Professor[2]
Benjamin G Druss, MD, MPH, Rosalynn Carter Chair in Mental Health[3]

1 Bryn Mawr Graduate School of Social Work and Social Research, Bryn Mawr, Pa, USA.
2 School of Social Policy and Practice, University of Pennsylvania, Philadelphia, Pa, USA.
3 Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, Ga, USA.
Correspondence: bdruss@emory.edu

REFERENCES

1 National Library of Medicine; National Institutes of Health. Errata, retraction, duplicate publication, comment, update and patient summary policy for MEDLINE. http://www.nlm.nih.gov/pubs/factsheets/errata.html (accessed Jun 2006).
2 Friedman PJ. Correcting the literature following fraudulent publication. JAMA 1990; 263: 1416-1419.
3 Pfeifer MP, Snodgrass GL. The continued use of retracted, invalid scientific literature. JAMA 1990; 263: 1420-1423.
4 Budd JM, Sievert M, Schultz TR. Phenomena of retraction: reasons for retraction and citations to publications. JAMA 1998; 280: 296-297.
5 Leape LL. Error in medicine. JAMA 1994; 272: 1851-1857.
6 International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication. Updated February 2006. http://www.icmje.org/ (accessed Jun 2006).
7 Garfield E, Welljams-Dorof A. The impact of fraudulent research on the scientific literature. The Stephen E Bruening case. JAMA 1990; 263: 1424-1426.
8 Whitely WP, Rennie D, Hafner AW. The scientific community's response to evidence of fraudulent publication. The Robert Slutsky case. JAMA 1994; 272: 170-173.
9 US Office of Science and Technology Policy (OSTP). Fact Sheet. Research misconduct — a new definition and new procedures for federal research agencies. 1999. http://clinton3.nara.gov/WH/EOP/OSTP/html/9910_20_2.html (accessed Jun 2006).
10 Thomson ISI. Journal citation reports. Science Edition. Philadelphia: Thomson, 2002.
11 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.
12 Eisner MD, Shiboski SC, Yelin EH, et al. Retraction. Am J Respir Crit Care Med 2000; 162 (3 Pt 1): 788.
13 Ricaurte GA, Yuan J, Hatzidimitriou G, et al. Retraction. Science 2003; 301: 1479.
14 Druss BG, Marcus SC. Growth and decentralization of the medical literature: implications for evidence-based medicine. J Med Libr Assoc 2005; 93: 499-501.
15 Greene PR. Independent re-analysis of identical data sets: implications for peer review. Fifth International Congress on Peer Review and Biomedical Publication; 2005 Sept 16-18; Chicago, Ill, USA.
16 Reason J. Human error. Cambridge, Mass: Cambridge University Press, 1990.
17 Druss B. Understanding errors in the research process. Lancet 2004; 363: 1230.
18 Garfield E. Journal editors owe readers timely action on retractions. The Scientist 1989; 3: 332-333. http://www.garfield.library.upenn.edu/essays/v14p332y1991.pdf (accessed Jun 2006).
19 Garfield E. How to avoid spreading error. Scientists must search for corrections. The Scientist 1987; 1: 255-257. http://www.garfield.library.upenn.edu/essays/v14p255y1991.pdf (accessed Jun 2006).