Effects of article retraction on citation and practice in medicine

By John M. Budd, Ph.D.
Associate Professor

School of Information Science and Learning Technologies
University of Missouri–Columbia
20 Rothwell
Columbia, Missouri 65211
libsjmb@showme.missouri.edu

MaryEllen Sievert, Ph.D.
Professor

School of Information Science and Learning Technologies and
Department of Health Management and Informatics
University of Missouri–Columbia
324 Clark Hall
Columbia, Missouri 65211

Tom R. Schultz, M.A.

Max Kade Center for German-American Studies
University of Kansas
Lawrence, Kansas 66045

Caryn Scoville

University of Missouri–Columbia
20 Rothwell
Columbia, Missouri 65211

At times, there are reasons for authors to make a formal statement of retraction of work they publish in biomedical journals. This study examines 235 retracted articles and looks at the reasons for these retractions and citations to the articles subsequent to retraction. The primary reasons for retraction are error of various kinds (such as problems with method or sample, including contamination of samples) and misconduct. The 235 articles are cited a total of 2,034 times after retraction. This set of citations can be divided into two groups: citations that appear in journals included in the Abridged Index Medicus and those that appear in other journals included in MEDLINE. While most of the citations in these two groups of journals can be categorized as “implicitly positive,” 275 make explicitly positive mention of retracted articles. The implications for continued citation for biomedical research and clinical practice are discussed.

INTRODUCTION

Every discipline relies on the integrity of its literature. The tacit assumption is that authors are diligent and honest, and that their contributions to journals and other outlets reflect that diligence and honesty. Every member of the discipline has a vested interest in the accuracy and authenticity of what is communicated. To be sure, the vast majority of the contributions to a discipline's literature maintain the high standards set for the legitimacy of research. The field of biomedicine is no different from others in relying on the efficacy of its primary communication vehicles. Researchers depend on their predecessors' zeal with regard to en-
suring appropriate controls on their work, conducting studies thoroughly and carefully, and reporting results completely. At times, however, one or more of these aims are not fulfilled. In some cases, unavoidable circumstances lead to contamination of samples or anomalies in data that result in erroneous findings being reported. Occasionally, carelessness has led to similar results. From time to time, researchers may fabricate or falsify data or findings and still report them as valid. Acknowledged problems of these kinds comprise a very small proportion of published works, and may be identified in the literature through retraction by one or more of the authors, by a journal’s editor, or by others. Errors that are recognized as such may also be retracted in the literature. A set of serious concerns remains: Who retracts publications? Why are they retracted? How are retracted publications received and used by subsequent researchers? These concerns form the foundation of the present study.

BACKGROUND

While error, as well as misconduct, can lead to retraction of published work, more attention has been paid to the more striking instances of misconduct and fraud. There have been some notorious cases of scientific misconduct in recent years. Broad and Wade [1], Kohn [2], LaFollette [3], and Whitley, Rennie, and Hafner [4] report on some of the more widely publicized cases and a collection of essays edited by Lock and Wells [5] addresses the matters of fraud and misconduct generally. Altman [6] reviews some of the most important aspects and outcomes of scientific misconduct. Lock [7] provides concise case studies of recent episodes of misconduct. Concern regarding misconduct has prompted the Public Health Service to establish an Office of Research Integrity (ORI)* some years ago. The ORI is charged with, among other things, tracking and overseeing investigation of allegations of scientific misconduct in biomedicine that may occur in intramural or extramural projects supported by the Public Health Service.

Some observers urge certain measures to detect misconduct and to prevent it in the future. For example, Hammerschmidt and Gross [8] suggest that textual analysis might be employed so as to examine circumstantial evidence (discrepancies between researchers’ resumes and published output; number and complexity of authors’ clinical studies; anomalies in the reported sequence of research activities, such as animal studies occurring after human studies; and publications reporting methods and techniques that have not yet been employed in clinical research). Others maintain that incidents of misconduct and fraud are seriously underreported (Stewart and Feder [9]). Still others are concerned about the social impact of misconduct. Woolf writes,

The damage caused by falsification is not related merely to its frequency. Even if cases of fraud are infrequent, fraud has an impact on the research of other working scientists, on the reliability of the published literature, and on public attitudes that are vital to the future support of research. Greater understanding of these effects would allow the scientific community to take steps to minimize the adverse consequences within science and to reassure the public that its trust is warranted [10].

While these concerns are legitimate and pressing, the biomedical literature is also affected by error that can render the reported results of research useless at best and dangerous at worst. The work done by Stewart and Feder [11] focuses on the occurrence of error. They have examined papers published by coauthors of John Darsee, who admitted to fraud in fabricating substantial bodies of research, and found that there have been numerous apparent errors in their published work. At times, the error can seem obvious; at other times, discovery of error requires a very close reading. At times, the authors themselves discover the error and may issue a retraction if they feel the error to be of sufficient magnitude and import that it must be acknowledged. While error inevitably occurs in research and sometimes finds its way into published work, it is less frequently examined than is misconduct. Awareness of the retraction and reasons for retraction may affect the frequency with which such papers are cited subsequent to retraction; this will be addressed shortly.

Whether because of misconduct or error, publications do occasionally get retracted. Researchers and clinicians must be able to identify readily those items that have been retracted so that they do not assume that the work is accurate or honest. In order to alert readers better, health sciences librarians have attempted to address the issue. Duggar et al. have conducted a survey of medical libraries and found that a majority of respondents have been involved in identifying retracted publications [12]. Cooper advocates that medical libraries search for retractions and errata and provide that information to the users of the library [13]. Bilko observes that the importance of identifying retracted publications extends beyond the local situation; document delivery and interlibrary services are also affected by the presence of the phenomenon of retraction [14]. Pfeifer and Snodgrass report the results of a survey of medical libraries and find that approximately 91% of retracted publications have not been marked in any way as invalid; moreover, they note that very few medical libraries have policies to handle alerting

* Information about and communications (including newsletters) from ORI may be viewed at http://www.os.dhhs.gov/phs/ori/oriHome.html.
readers of retraction [15]. A more recent study by Hughes indicates that more libraries (32% of responding institutions) have policies and procedures for notifying users about retractions, and another 9% call attention to retractions even though they have no formal policies on the matter [16]. Some have addressed the matter of continued citation of retracted items. Pfeifer and Snodgrass found that citation continued and that, while there was diminished citation activity in journals published in the United States (from preretraction to postretraction), the number of citations was still high and that postretraction citations actually exceeded preretraction citations in journals published in Europe [17]. Garfield and Welljams-Dorof looked at citations to the work of Stephen Breuning, who admitted to fraud, and found that activity had greatly diminished and that many of the citations were negative. They concluded,

The most interesting observation to emerge from this single case study is that the scientific literature seems to purge itself of articles that are known or even suspected to be fraudulent. The annual distribution of non-self-citations indicates that authors shun falsified research once it is publicly exposed [18].

This conclusion is not borne out in the study by Kochan and Budd. They examined citations to works by John Darsee, who also admitted to fraud, and found a substantial number of postretraction citations, the vast majority of which were implicitly or explicitly positive [19].

Based on the available background research, a set of assumptions inform this study. First, authors are assumed to be the ones doing the retracting, because they would be in the position to be aware of errors or misconduct. Second, most retractions are assumed to occur because of scientific misconduct or unavoidable error. Third, because the problem or problems would be of such magnitude, an entire article was assumed to not be regarded as a valid scientific article.

METHODS

In order to obtain a body of data to analyze, the MEDLINE database (from 1966 through 1996) was searched, using the publication type 'retraction of publication.' The search yielded 198 statements of retraction that appeared in the journal literature. Figure 1 illustrates the appearance of a retraction in the MEDLINE database (this record is retrieved from the OVID system). Some of these statements retracted more than one article; a total of 235 articles were retracted. Figure 2 provides an example of the retraction of multiple articles.

The next task was to classify the articles according to: who retracted the publication; what content was retracted; why the article (or portion of the article) was retracted; and how long after publication the retraction occurred.

In order to determine citing activity, Science Citation Index was searched so as to identify all citations to each of the retracted publications. Pre- and postretraction citations were identified separately in order to determine to what extent each paper was cited after it had been retracted. A one-year period was inserted before a citation was considered as being postretraction. This period was given to allow for indexing to be in place so that searches of the MEDLINE database would have ready access to the retraction statement. Also, the one-year period compensates for publication lag; that is, if something was in press at the time a retraction statement was published, then that article could, in good faith, contain citations to retracted items. The 235 retracted articles received 2,034 postretraction citations as of January 1, 1997. For the purpose of analysis, these citations were divided into two groups. One focus for this study was on postretraction citations appearing in journals indexed in the Abridged Index Medicus (AIM) because, as Roper and Boorkman noted, AIM was designed to be used by practicing physicians and included important clinical journals [20]. This set of postretraction citations would appear in those journals that were most likely to have an impact on clinical work and treatment of patients. The other focus was on ci-
tations in articles in non-AIM journals. For both types of journals the postretraction citations could be divided into three categories: the citing article acknowledged the retraction; the citing article explicitly cited the retracted article as presenting valid research (usually by making specific reference to the paper); or the citing article implicitly cited the retracted article as valid (for instance, by including reference to the retracted article along with other articles). A chi-square test could be employed to determine if there were differences in citation patterns with regard to positive or negative citation between the citations in the two groups of journals.

In addition to noting the category of citation (just mentioned), both the kind of citing publication (letter, review article, or article) and the place (section such as introduction, methods, etc.) in the citing article (introduction, methods, discussion, results, or conclusions) the citation occurs can be accounted for.

RETRACTED ARTICLES

As is mentioned above, the authors were assumed to be retracting the articles at the outset. As it turned out, one or more of the authors retracted 190 of the 235 articles; 45 were retracted by others, including investigating committees or deans, editors, or legal counsels. All of the paper was also assumed to be retracted. In 200 instances, the entire paper was retracted; 35 cases part, but not all, of the paper was retracted. For example, Clark, Fletcher, and Petersen stated:

We wish to retract this sentence from the case report section: 'Although they were willing to abort as many fetuses as necessary prior to conceiving one which met these specifications, this approach posed ethical problems for our staff.' We wish to substitute instead: 'The possibility of terminating one or more HLA incompatible fetuses posed ethical problems for our staff' [21].

As stated above, one purpose of the study was to determine the length of time between publication of an article and its retraction. The mean time from publication to retraction was twenty-eight months. This mean included the retractions by one author who retracted four articles, all ten years after their publication. Controlling for this anomalous case yielded a mean time from publication to retraction of 25.8 months. The range was from 2 to 197 months.

The question of why articles were retracted was also a purpose of the preliminary investigation and articles were assumed to be retracted because of misconduct or honest error. In ninety-one articles, retraction occurred because of some kind of error. The categories of error that could be identified were: error in the methods or analysis (N = 23); problems with the data (N = 37); and problems with the sample, such as contamination (N = 31). A total of eighty-six papers were retracted because of misconduct or presumed misconduct. Retraction was classified as being due to misconduct if the statement of retraction clearly admitted to wrongdoing on the part of one or more of the authors. Presumed misconduct referred to those instances where one or more of the authors raised serious questions about the efficacy of the work done by other authors. For example, one retraction statement read, "This article is retracted by its first author (T. Nitta) in agreement with his co-authors, due to several important inaccuracies. Dr. Steinman's name was used without his knowledge or permission. He discovered the inaccuracies and made them known to the Editor-in-Chief" [22]. An additional thirty-eight articles were retracted because the authors could not replicate the results. These articles were categorized separately because it was not possible to determine if the results were not replicated because of unavoidable error or because of some wrongdoing. The remaining twenty articles were retracted for other, unclassifiable reasons. In three instances, the retractors were unclear as to reason; in three cases, idiosyncratic reasons were given; in fourteen cases, no reason was given. Table 1 illustrates the breakdown of retractions by reason.

CITATIONS IN AIM JOURNALS

The 235 articles received a total of 2,034 postretraction citations. Of these citations, only 299 appear in journals indexed in AIM. The analyses of these items have been reported on by the authors [23]; summary of those findings are repeated here. The majority of the citations in the AIM journals were found in articles reporting research (N = 277). Only fourteen citations appeared in letters, and five of those made reference to the retraction. Review articles accounted for eight citations, and only one of those acknowledged the retraction. This last finding was noteworthy because review articles should be complete in their reference to the literature of a particular subject, including statements of retraction, yet in seven out of eight instances no mention was made of the retraction. Also, review articles are generally highly regarded in the profession and are frequently recommended to medical students as a means of becoming familiar with a topic.

Of the 299 postretraction citations appearing in AIM journals, 19 acknowledged the retraction in some way. As was just mentioned, 5 of these acknowledgments

<table>
<thead>
<tr>
<th>Error</th>
<th>Misconduct</th>
<th>Could not replicate</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>86</td>
<td>38</td>
<td>20</td>
</tr>
<tr>
<td>36.7%</td>
<td>36.6%</td>
<td>16.2%</td>
<td>8.5%</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>Acknowledge retraction</th>
<th>Explicitly positive</th>
<th>Implicitly positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIM citations (N = 299)</td>
<td>6.4% (N = 19)</td>
<td>5.7% (N = 18)</td>
</tr>
<tr>
<td>Non-AIM citations (N = 1,594)</td>
<td>7.7% (N = 123)</td>
<td>16.1% (N = 257)</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Acknowledge retraction</th>
<th>Explicitly positive</th>
<th>Implicitly positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>7.7%</td>
<td>16.1%</td>
</tr>
<tr>
<td>257</td>
<td>16.1%</td>
<td>76.1%</td>
</tr>
</tbody>
</table>

Effects of article retraction

able citations were usually brief mentions or inclusions of bibliographic references; these citing article did not question the validity of the retracted work. For instance, one citing article stated, “Another possible mechanism is myocardial stunning, which has been shown to be present for several hours and days after repeated and brief episodes of ischemia [followed by references to three articles, including a retracted article]” [27]. Table 3 illustrates the distribution of citations in non-AIM journals, including percentages. The overall distribution indicates that over 16% of the citations are explicitly favorable. Moreover, if only the citations to those retracted publications that are classified as being due to misconduct or presumed misconduct (N = 346), then it becomes evident that 27.2% are explicitly positive (N = 94). Only twenty of those citations (5.8%) make clear mention of the reason for retraction. By way of comparison, among the AIM journals only eight out of a total of eighty citations (10%, as opposed to the 27.2% for non-AIM journals) are to publications classified as due to misconduct, and eleven citations (13.8%, compared with the 5.8% for non-AIM journals) mention the reason for retraction. The difference in positive citations may be due, in part, to the selectivity of AIM; the journals included in that source are generally the more prestigious and widely read. Articles that are more critical of background work may be accepted by AIM journals.

The division of analysis into the AIM and non-AIM journals offers an opportunity for further analysis. As previously mentioned, the rationale for dividing the investigation into these two groups centers on the centrality of the journals included in AIM to the clinical practice of medicine. This rationale does not mean that the non-AIM journals are not read or do not have an impact on research and practice. A chi-square test can be employed to determine if the two distributions represent a good fit with one another. The computed value of chi-square is 24.08, which indicates a significant difference between the AIM and non-AIM journals (P < 0.05).

Another matter of concern in this study is where in the citing articles do citations to retracted publications appear. Most research reports are divided into sections usually labeled introduction, methods, discussion, results, and conclusions. Appearance of citations to retracted publications in the methods, results, or conclusion sections are likely to indicate a more substantive
use of the publication in question. In fact, the majority of citations appear in the introduction or discussion. A total of 117 citations are made in the introduction; 153 appear in the discussion. On the other hand, 47 appear in the methods section, 6 in the results, and 2 in the conclusion. The total here is 299; citation may occur in more than one section of a paper.

As is true with citations in AIM journals, most citations in research articles appearing in non-AIM journals occur in the introduction and discussion sections of the articles (introduction, N = 421; discussion, N = 272). However, some citing authors include retracted material in more substantive sections of their work. For instance, 107 citations occur in the methods section, 68 in the results section, and 4 in the conclusion. A number of citing articles are not divided into specific section; citations appearing in these articles are not counted here. Such citation practice indicates that, to some extent, the retracted work represents formative or evaluative information for the citing authors. The remaining citations occur in source documents other than research articles. Among the citations in items other than research articles, a total of 234 citations appear in review articles, so, in a sense, these citations could be categorized as appearing in a discussion of the topic in question. The final 33 citations that could be analyzed appear in letters.

DISCUSSION

As indicated above, this study has operated under the assumptions that the authors would be retracting the articles, that retractions would be made because of misconduct or error, and that the entire work would be retracted. The second of the stated expectations has proved to be somewhat problematic (the reasons for retraction proved to be more varied than expected); the first and third assumptions have been shown to be, with a few exceptions, supported. Additionally, the assumption that the retracted articles would continue to be cited as valid was supported. The results of this study strongly indicate that retraction of a publication, even though the retraction may be visible in the journal and is clearly noted in the MEDLINE database, does not ensure that all subsequent researchers will be alerted to the retraction and will cease making reference to the retracted work. What the cause of retraction was seemed to matter little; citations might well continue to any retracted article.

While 263 of the citing articles in AIM journals and 1,214 of the citations in non-AIM journals in this study embody implicitly positive citation to retracted publications, and while these citations tend to appear in the introduction or discussion sections of articles, the citations still ensure that the retracted articles continue to appear in citation indexes and that they may be retrieved by readers of the citing article. (It should be noted that, of all citations analyzed, 154 citations appear in the methods section, 74 in the results, and 6 in the conclusion.) If researchers come upon one of the citing papers and find the work done there of use, then they may turn to cited works and incorporate them into their work as though the work were valid. Such researchers, who do not retrieve the information through a formally structured MEDLINE search, may be unaware of the retraction.

It should be pointed out that biomedical science tends to be self correcting; that is, work that is not replicable due to error or misconduct is usually dismissed in time. However, there may be a great deal of time, effort, and money spent in discovering that some research is not useful. If erroneous or fraudulent work lives on in the literature, the amount of time, effort, and money to correct work may be even greater. It should also be noted that this study focuses on work that has been acknowledged to be based on error or misconduct. The question remains as to how much erroneous or fraudulent work goes undetected or unacknowledged. That is a larger question that should be addressed by the biomedical community. The evidence provided by this study suggests that it is a serious question with profound implications both for research and for clinical practice.

In addition to retractions of published work, some other anomalies can occur in the biomedical literature and can pose problems that are similar to those of retractions. Future work should address the occurrence and nature of errata, corrections to published articles, and duplicate publications. The population of these publication types is large and includes, in some instances, major corrections or notations of error in published works. Examination of these phenomena, in conjunction with the results presented here, will offer a clearer picture of the complex dynamics of communication in a discipline that includes research and clinical work that can have an impact on the majority of the world’s population.

REFERENCES

11. STEWART, op. cit.

Received April 1999; accepted May 1999