Phenomena of Retraction

Reasons for Retraction and Citations to the Publications

John M. Budd, PhD; MaryEllen Sievert, PhD; Tom R. Schultz, MA

Context.—This study examined the impact of retracted articles on biomedical communication.

Objective.—To examine publications identified in the biomedical literature as having been retracted, to ascertain why and by whom the publications were retracted and to what extent citations of later-retracted articles continue to be incorporated in subsequent work.

Design.—A search of MEDLINE from 1966 through August 1997 for articles that had been retracted.

Main Outcome Measures.—Characteristics of retractions and citations to articles after retraction.

Results.—A total of 235 articles had been retracted. Error was acknowledged in relation to 91 articles; results could not be replicated in 38; misconduct was evident in 86; and no clear reason was given in 20. Of the 235 articles, 190 were retracted by some or all of the authors; 45 were retracted by a person or organization other than the author(s). The 235 retracted articles were cited 2034 times after the retraction notice. Examination of 299 of those citations reveals that in only 19 instances was the retraction noted; the remaining 280 citations treated the retracted article either explicitly \( (n = 17) \) or implicitly \( (n = 263) \) as though it were valid research.

Conclusion.—Retracted articles continue to be cited as valid work in the biomedical literature after publication of the retraction; these citations signal potential problems for biomedical science.

VARIOUS FACTORS can lead to the retraction of a publication. There have been some notorious cases of scientific misconduct in recent years. LaFollette\(^1\) and Whitley et al\(^2\) report on some of the more widely publicized cases, and a collection of essays edited by Lock and Wells\(^3\) addresses generally the matters of fraud and misconduct. 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\(^4\) focuses on the occurrence of error. Awareness of the retraction and reasons for retraction might affect the frequency with which such articles are cited subsequent to retraction.\(^5\)

Based on the available background research, a set of expectations can be stated: the authors are the ones doing the retracting; most retractions occur because of scientific misconduct or unavoidable error; the entire article should be regarded as an invalid scientific article; and retracted publications continue to be cited after retraction.

METHODS

The data for analysis came from a MEDLINE search (1966-August 1997) that used the publication type “retraction of publication.” This strategy includes 2 limitations: an article must be formally retracted to be assigned this publication type, and indexers must recognize the retraction and assign the publication type. Further, this study focuses entirely on retraction, so corrections and errata are not included. The search yielded 235 retracted articles. This population is a very small subset of the MEDLINE database, but it does include all publications formally identified as retractions. The next task was to classify the articles according to the following characteristics: 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.

To determine citing activity, Science Citation Index was searched to identify all citations to each of the retracted publications. A 1-year period after publication of the retraction was inserted before a citation was considered as postretraction to allow for indexing of the retraction to be in place. Searchers of the MEDLINE database would then have ready access to the retraction statement. Also, the 1-year period compensates for publication lag; that is, if a manuscript is in press at the time a retraction statement is published, then that article could, in good faith, contain citations to retracted items. This study focuses on postretraction citations appearing in journals indexed in the Abridged Index Medicus (AIM), since that source contains the most clinically relevant journals. The postretraction citations were divided into 3 categories: the citing article acknowledged the retraction, the citing article explicitly cited the retracted article as presenting valid research, or the citing article implicitly cited the retracted article as valid. In
addition to noting the category of citation, the kind of citing publication (letter, review article, or article) and where in the citing article (introduction, methods, results, discussion, or conclusions) the citation occurs were recorded.

RESULTS

One element of the phenomenon of retraction addressed was the length of time between publication of an article and its retraction. The mean time from publication to retraction was 28 months. This mean includes the retractions by 1 author of 4 articles 10 years after their publication. Controlling for this anomalous case yields a mean time from publication to retraction of 25.8 months (range, 2-197 months).

One or more of the authors retracted 190 of the 235 articles; 45 were retracted by others, including institutional investigating committees or deans, journal editors, or legal counselors. Retraction of 91 articles occurred because of some kind of error. The categories of error identified were the following: error in the methods or analysis in 23, problems with the data in 37, and problems with the sample (such as contamination) in 31. A total of 86 articles were retracted because of misconduct or presumed misconduct. Retraction was classified as due to misconduct only if the statement of retraction clearly admits to wrongdoing on the part of one or more of the authors. Presumed misconduct refers to those instances where one or more of the authors raises serious questions about the efficacy of the work done by other authors. An additional 38 articles were retracted because the author(s) could not replicate the results. These were categorized separately because it is not possible to determine if the results were not replicable because of unavoidable error or because of some wrongdoing. The remaining 20 articles were retracted for other unclassifiable reasons. In 3 instances the retractors were unclear as to reason, in 3 cases idiosyncratic reasons were given, and in 14 cases no reason was given. In 200 instances the entire article was retracted; in 35 cases part, but not all, of the article was retracted.

The 235 articles received a total of 2034 postretraction citations. Of these citations, 299 appeared in journals indexed in AIM. The majority of the citations were found in articles reporting research or clinical practice (n = 277). Only 14 citations appeared in letters, and 5 of those letters referred to the retraction. Review articles accounted for 8 citations, and only 1 of those acknowledged the retraction. Of the 299 postretraction citations appearing in AIM journals, 19 acknowledged the retraction in some way. Given that 5 of these acknowledgments appeared in letters and 1 appeared in a review article, only 13 articles reporting research made specific mention of the retraction. Of the remaining citations, 17 explicitly treated the retracted article as valid, usually by naming the authors of the article or mentioning specific elements of their findings or methods. The remaining 263 citations include implicit approval of the retracted work, usually in the form of brief mention or bibliographic reference in a passage that in no way questions the validity of the research.

COMMENT

The second of the stated expectations proved to be somewhat problematic; the first, third, and fourth were shown to be, with a few exceptions, 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. It seems to matter little if the cause of retraction is error or misconduct; citations to any retracted article may well continue.

While 263 of the citing articles in this study embody implicit positive citation to retracted publications, and while these citations tend to appear in the introduction (n = 117) or discussion (n = 153) 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 47 citations appear in the methods section, 6 in the results, and 2 in the conclusion.) If a researcher comes upon one of the citing articles and finds the work done there of use, then he or she may turn to cited works and incorporate them into his or her work as though the work were valid. Such a researcher, who does not retrieve the information from 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 because of 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 spent in discovering that some research is not useful. If 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.

References


We agree with Berman that there may often be honest differences of opinion among informed experts about the value of new treatments. As stated in our article, our study cannot establish whether current practices are right or wrong. However, the wide gap between the recommendations of physicians and the approvals of insurers in the cases under study underscores a potentially far-reaching problem in the US health care system. We believe that insurers often make very careful and fair analyses of information in arriving at coverage decisions. Nevertheless, greater dialogue between insurers and physicians is needed to minimize discrepancies in coverage and to provide optimal patient care.

Dr Yaes asks about consensus among physicians for the cases studied and about the value of a treatment such as GH, which is used for a non–life-threatening condition. We focused on GH therapy as a key example of emerging treatments that are semielective, relate at least in part to quality of life, and for which consensus about optimal utilization is lacking. As we pointed out in our article, many emerging therapies share similar characteristics, e.g., treatments for infertility, impotence, and aging. For GH therapy, lack of consensus exists for severe cases of idiopathic short stature, whereas there is much more consensus among physicians for GH treatment of children with Turner syndrome and renal failure. Nevertheless, we found significant discrepancies between physician recommendations and insurer coverage for Turner syndrome and renal failure, as for idiopathic short stature.

Ms Tesch and Dr Yaes differ widely in their viewpoints and underscore the fact that the value of treatments for extreme short stature (or other conditions that are semielective and address, to some extent, quality of life) is based, in part, on the degree to which the underlying condition is perceived as representing a form of morbidity. As more such treatments emerge, clinicians will increasingly be confronted with difficult questions about their use. Our study illustrates the real-life difficulties in delivering optimal and equitable health care for nonemergency conditions. We believe that the discrepancy between physician treatment recommendations and insurance coverage constitutes an important challenge to health care delivery and that patients deserve increased and constructive dialogue between physicians and insurers in arriving at coverage decisions.

Leona Cuttler, MD
Beth S. Finkelstein, PhD
J. B. Silvers, PhD
Duncan Neuhauser, PhD
Ursula Marrero, MSSA
Case Western Reserve University
Cleveland, Ohio

In Reply.—We agree with Dr Berman that insurers vary substantially in the procedures they use to develop and implement policy regarding coverage for new technologies (and other treatments) and in the values they apply in making their decisions. We also agree that many insurers make serious efforts to address the 3 conditions we believe are necessary to create a public sense of legitimacy and fairness for limit-setting policies: (1) articulate the rationale behind their policies, (2) make the rationale public, and (3) provide opportunities for clinicians and members to appeal policies and specific decisions. However, even in the outstanding programs we were privileged to study in our research on policymaking, we and the program leaders agreed there was room for significant improvement.

Ms Tesch's letter on behalf of the Turner's Syndrome Society provides an example of well-conceived advocacy. The fact that GH has been recognized by the Food and Drug Administration as effective in the treatment of Turner syndrome puts a burden of explanation onto an insurer who chooses not to cover it. In asking whether decisions to deny coverage reflect "reasonable cost-benefit analysis or an inappropriate denial of care," Tesch raises the key policy question in useful terms. In a poor country, denial would be easily defendable in terms of competing priorities, but in a country as wealthy as the United States, the case for denying coverage is more difficult to justify.

Dr Yaes notes the degree of controversy that surrounds use of GH to treat idiopathic short stature. We disagree with his view that the controversy makes GH a bad focus of study. The controversy actually makes GH an excellent "biopsy" of policymaking and underscores the importance of articulating the rationale for noncoverage. If denial of coverage for GH in idiopathic short stature is based on doubt about its efficacy, proof of efficacy would refute the objection. If the denial is based on fear of adverse effects, an informed patient and family might argue that they understand and choose to accept the risks. If the denial is based on regarding any benefits as "purely cosmetic," then the issue is one of our understanding of short stature and the goals of medicine. Yaes' comments are a contribution to the kind of deliberative process on which coverage policies should be based.

James E. Sabin, MD
Harvard Pilgrim Health Care
and Harvard Medical School
Norman Daniels, PhD
Tufts University
Boston, Mass

CORRECTIONS

Error in Reference List.—In the article entitled "Reporting of Randomized Clinical Trial Descriptors and Use of Structured Abstracts," published in the July 15, 1998, Peer Review theme issue of THE JOURNAL (1998;280:269-272), an error was made in the reference list. On page 272, the second reference 9 should be deleted.

Incorrect Unit of Measure.—In the Letter by Rim et al entitled "Relationship of Dietary Folate and Vitamin B6 With Coronary Heart Disease in Women" published in the August 5, 1998, issue of THE JOURNAL (1998;280:418-419), there was an incorrect unit of measure listed. On page 419, the last sentence of the second full paragraph reads, "The concern raised by Herbert about masking anemia due to vitamin B12 deficiency does not apply to the usual amount of folate (400 g/d) consumed by women in our cohort as part of a multiple vitamin." The amount of folate should have been 400 µg/d.