

SCIENTIFIC INTEGRITY

Responding to Possible Plagiarism

Tara C. Long,¹ Mounir Errami,² Angela C. George,¹ Zhaohui Sun,² Harold R. Garner^{1,2*}

The peer-review process is the best mechanism to ensure the high quality of scientific publications. However, recent studies have demonstrated that the lack of well-defined publication standards, compounded by publication process failures (1), has resulted in the inadvertent publication of several duplicated and plagiarized articles.

The increasing availability of scientific literature on the World Wide Web has proven to be a double-edged sword, allowing plagiarism to be more easily committed, while simultaneously enabling its simple detection through the use of automated software. Unsurprisingly, various publishing groups are now taking steps to reinforce their publication policies to counter the fraudulent acts of a few (2). There are now dozens of commercial and free tools available for the detection of plagiarism. Perhaps the most popular programs are iParadigm's "iThenticate" (<http://ithenticate.com/>) and TurnItIn's originality checking (<http://turnitin.com/>), which recently partnered with CrossRef (<http://www.crossref.org/>) to create CrossCheck, a new service for verifying the originality of scholarly content. However, the content searched by this program spans only a small sampling of journals indexed by MEDLINE. Others include EVE2, OrCheck, CopyCheck, and WordCHECK, to name a few.

We recently introduced an automated process to identify highly similar citations in MEDLINE (3, 4). Our detection of duplicates relies heavily on human inspection in conjunction with computational tools including eTBLAST (5, 6) and Déjà vu, a publicly available database (7, 8). As of 20 February 2009, there were 9120 entries in Déjà vu with high levels of citation similarity and no overlapping authors. Thus far, full-text analysis has led to the identification of 212 pairs of articles with signs of potential plagiarism. The average

text similarity between an original article and its duplicate was 86.2%, and the average number of shared references was 73.1%. However, only 47 (22.2%) duplicates cited the original article as a reference. Further, 71.4% of the manuscript pairs shared at least one highly similar or identical table or figure. Of the 212 duplicates, 42% also contained incorrect calculations, data inconsistencies, and reproduced or manipulated photographs.

There has been a paucity of literature examining the reactions of stakeholders (both victims and perpetrators) when confronted with evidence of possible misconduct. Studying these reactions may help to illuminate the reasons for such misconduct and might provide a way for the scientific community to prevent such activity in the future. Therefore, we merged data from previous studies (3) with additional information based on our personal communications with authors and journal editors directly associated with 163 of these cases of potential plagiarism.

A questionnaire (see table S1) was composed, supplemented with annotated electronic copies of both manuscripts, and sent via e-mail to the authors and editors of the earlier and later manuscripts.

From the 163 sets of questionnaires sent, we received a reply in 144 cases (88.3%). Anonymity was guaranteed to all respondents. The reactions by the respondents were intense and diverse, and although it is difficult to quantify the various responses, a general picture can be painted. Before receiving the questionnaire, 93% of the original authors were not aware of the duplicate's existence. The majority of these responses were appreciative in nature. The responses from duplicate authors were more varied; of the 60 replies, 28% denied any wrongdoing, 35% admitted to having borrowed previously published material (and were generally apologetic for having done so), and 22% were from coauthors claiming no involvement in the writing of the manuscript. An additional 17% claimed they were unaware that their names appeared on the article in question. The journal editors primarily confirmed receipt and addressed issues involving policies and potential actions. Excerpts from statements made by authors and editors

Documenting reactions from authors and journal editors to plagiarism may help others address the problem.

illustrate the many possible perspectives in response to evidence of possible plagiarism. Table 1 provides a sampling of these responses, with an expanded list available in tables S2 to S5.

Although the goal of the questionnaire was merely to solicit information, the very act of sending it appeared to trigger further action by journals in many cases. Editors have launched 83 internal investigations thus far, 46 of which have, according to the editors of the journals, led to eventual retraction of the duplicate article. It is unclear what defines a "retraction," however, because many editors only stated that a comment would be published in their journal, or that the article would simply be removed from the journal's Web site. Unfortunately, these actions do not propagate back to MEDLINE unless an explicit request is made by the journal; therefore, researchers and clinicians may never become aware of an article's retracted status.

To assess how articles of this nature affect the scientific community, we recorded the impact factors for each journal in which the 212 articles and their duplicates were published using the Thomson Scientific Journal Citation Reports feature (9). A large portion of the duplicates were published in low-profile journals; thus, impact factors were available for only 199 of the 285 different journals. The impact factors of journals publishing original articles were significantly higher ($P < 0.001$), averaging 3.87 and spanning 0.147 to 52.589, than those of the journals publishing duplicate articles, averaging 1.6 and spanning 0.272 to 6.25.

Utilizing the ISI Web of Knowledge to determine how many times each article had been cited (10), we found that original publications were cited 28 times on average, whereas their corresponding duplicates were cited only twice. Although the original articles are older and have thus had more exposure, in 10 of the pairs, the duplicate article was cited at least as often as the original publication. This may be because scientists rely heavily on finding information through PubMed searches which, by default, return more recent articles first, ensuring that a plagiarized article will always appear higher on a list of search

¹McDermott Center for Human Growth and Development, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9185, USA. ²Division of Translational Research, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9185, USA.

*Author for correspondence. E-mail: harold.garner@utsouthwestern.edu

results than its original counterpart. As a result, citations that would have otherwise gone to an original publication are instead diverted to a plagiarized one.

Of the 175 journal editors with whom we communicated, 11 admitted they had never personally dealt with a potentially plagiarized manuscript and were unsure how to proceed.



Sampling of responses from authors and editors

Authors of earlier article

"I have been a research scientist for more than 50 years, and this is the first time I've ever experienced such a blatant case of plagiarism. It sure was an eye-opener!"

"I have no statement. I cannot prove that this is plagiarism. Even if it is, what can be done?"

"[My] major concern is that false data will lead to changes in surgical practice regarding procedures."

"We were very sorry and somewhat surprised when we found their article. I don't want to accept them as scientists."

Editors of journal publishing earlier article

"It's my understanding that copying someone else's description virtually word-for-word, as these authors have done, is considered a compliment to the person whose words were copied."

The two articles "are the same patients, the figures are the same, and the writing is blatant plagiarism. One of these papers is a false publication. We cannot let this one go unaddressed."

"We were not aware of this duplicate publication, and would not have given permission for this, as it clearly violates copyright."

"I have been Editor for 14+ years and this is the first time this issue has been raised."

"It is clear that the subsequent author frankly, fraudulently used identical data ... in writing the second article. There is no way under the stars that we could have picked that up ourselves."

Authors of later article

"I would like to offer my apology to the authors of the original paper for not seeking the permission for using some part of their paper. I was not aware of the fact I am required to take such permission."

"There are probably only 'x' amount of word combinations that could lead to 'y' amount of statements. ... I have no idea why the pieces are similar, except that I am sure I do not have a good enough memory—and it is certainly not photographic—to have allowed me to have 'copied' his piece.... I did in fact review [the earlier article] for whatever journal it was published in."

"I know my careless mistake resulted in a severe ethical issue. I am really disappointed with myself as a researcher."

"It was a joke, a bad game, an unconscious bet between friends, 10 years ago that such things ... happened. I deeply regret."

"I was not involved in this article. I have no idea why my name is included."

Editors of journal publishing later article

"Looks like [the author of the later article] did it again in 2001. This example is a bit more embarrassing because the author of the original paper is [the] editor of the journal where [the author of the later article] published the copied work. Looks like we will have to publish two retractions."

"Believe me, the data in any paper is the responsibility of the authors and not the journal."

"I really appreciate your work and your e-mail has promoted us to exercise more strict control over duplicate publication."

"There can be no doubt that this is willful and deliberate plagiarism. Like the chance of monkeys typing out the works of Shakespeare, it would be incredible that the similarities could arise by chance."

"The news has taken us by surprise and a sense of deep concern. We are calling an emergency meeting of the editorial board to discuss the matter. [Our journal] deeply condemns the act and we stand firm to take necessary actions against the authors."

The majority of these editors showed deep concern and were open to any helpful suggestions or recommendations we could offer, at which point we directed them to the Office of Research Integrity's guidance document for editors on *Managing Allegations of Scientific Misconduct (11)*. In spite of this concern, nearly half of all the duplications brought to light by our questionnaires have

received no action. In fact, on 12 separate occasions, editors specifically indicated that cases involving their journal would not be reviewed. This variation in feedback reveals a great deal about the attitudes and motivations of scientists around the globe, including why some journal editors do not pursue obvious cases of duplication. Some apparently do not want to deal with the added stress of conducting a thorough investigation. Others feel it may bring bad publicity or reflect poorly on their journal's review process.

While there will always be a need for authoritative oversight, the responsibility for research integrity ultimately lies in the hands of the scientific community. Educators and advisors must ensure that the students they mentor understand the importance of scientific integrity. Authors must all commit to both the novelty and accuracy of the work they report. Volunteers who agree to provide peer review must accept the responsibility of an informed, thorough, and conscientious review. Finally, journal editors, many of whom are distinguished scientists themselves, must not merely trust in, but also verify the originality of the manuscripts they publish.

References and Notes

1. L. Gollogly, H. Momen, *Rev. Saude Publica* **40**, 24 (2006).
2. C. White, *BMJ* **336**, 797 (2008).
3. M. Errami *et al.*, *Bioinformatics* **24**, 243 (2008).
4. Materials and methods are available as supporting material on *Science Online*.
5. J. Lewis, S. Ossowski, J. Hicks, M. Errami, H. R. Garner, *Bioinformatics* **22**, 2298 (2006).
6. M. Errami, J. D. Wren, J. M. Hicks, H. R. Garner, *Nucleic Acids Res.* **35**, W12 (2007).
7. M. Errami, H. Garner, *Nature* **451**, 397 (2008).
8. M. Errami, Z. Sun, T. C. Long, A. C. George, H. R. Garner, *Nucleic Acids Res.* **37**, D921 (2009).
9. Journal Citation Reports, ISI Web of Knowledge (Thomson Reuters, Philadelphia, 2008); http://isiwebofknowledge.com/products_tools/analytical/jcr/.
10. ISI Web of Knowledge (Thomson Reuters, Philadelphia, 2008); <http://isiknowledge.com>.
11. Office of Research Integrity (ORI), *Managing Allegations of Scientific Misconduct: A Guidance Document for Editors* (ORI, U.S. Department of Health and Human Services, Rockville, MD, 2000); http://ori.dhhs.gov/documents/masm_2000.pdf.
12. We thank D. Trusty for computer administrative support; J. Loadsman as a substantial contributing curator; W. Fisher for useful comments, discussions, and manuscript editing; D. Wu and W. Fisher for assistance in obtaining full text articles; L. Gunn for administrative assistance; and the numerous Déjà vu users who have reported inaccuracies or have alerted us to questionable publications. This work was funded by NIH grant 5R01LM009758-02, the Hudson Foundation, and the P. O'B. Montgomery Distinguished Chair.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1293/DC1

10.1126/science.1167408