

Letters

Letter to the Editor



Do Open-Access Articles Really have a Greater Research Impact?

In her *C&RL* article, “Do Open-Access Articles Have a Greater Research Impact,” Kristin Antelman concluded that “[articles] have a greater impact as measured by citations ... when their authors make them freely available on the Internet.”¹ Her definition of Open Access was “any freely available full-text version (including drafts, preprints, and postprints).”² Because it was impossible to do a controlled experiment, Antelman compared the citedness of articles that she could find freely online with articles that she could not. This difference, she concluded, was the result of Open Access.

Antelman makes a causal argument – that OA publication *causes* more citations – and in the absence of other explanatory models, this is a very easy conclusion to make. Yet in order to demonstrate causation in science, three conditions must be met: 1) covariation; 2) temporal order (the hypothesized cause precedes the effect); and 3) the rejection of alternative explanations. While there is evidence of covariation and temporal order, I will argue that two alternative explanations are just as plausible and likely for explaining increased impact.

Article Duplication as Cause

From 1975 through 2003, Emerald (formally MCB University Press) engaged in substantial article duplication, where hundreds of articles were republished in different journals with no indication of their original source.^{3, 4} This dataset provides a natural controlled experiment to test whether the simple act of article duplication may explain increased citation impact. It is a controlled experiment in that one can compare the citations

received by the *very same article* published in two or more journals. Results indicate, not surprisingly, that those articles that were published in two or more journals received more citations than a random sample of articles published in only one journal.⁵ While this example does not involve openly accessible drafts, preprints and postprints, the plausible explanation is that multiple publishing increases the visibility of articles to diverse communities, thus increasing the chance an article will be read and cited. Viewed under the same rubric, *OA as article duplication* can explain Antelman’s findings.

In 2001, Kent Anderson and others published a citation study of the journal *Pediatrics*.⁶ Using three years of data, they compared articles printed in the journal and available online by subscription with other articles appearing only in a free online addition. The authors’ main findings suggested that despite wider potential audience for articles published freely online, articles appearing in print received, on average, about three more citations. While these findings obviously do not support Antelman’s hypothesis, they are consistent within the explanation of *article duplication as cause*. Those articles published in print and online (by subscription) may have provided increased visibility to potential readers than articles available freely from the *Pediatrics* web site.

Self-Promotion as Cause

In 2005, Jonathan Wren, a bioinformatics researcher at the University of Oklahoma conducted a massive automated study of the availability of author reprints on the public web.⁷ He reported two main

conclusions: that articles available freely online yielded more citations; and that there was a high degree of association between high-prestige journals and frequency of author reprints. Journals with high Impact Factors (*New England Journal of Medicine*, *Nature*, *Science*, and *Cell*) were associated with a higher degree of author republishing than lower-impact journals. Wren went further to discuss possible causes of this difference and briefly discusses a “trophy effect – the desire for researchers to display their accomplishments – which would explain why high impact publications are more common online.”⁸ This is consistent with Antelman’s findings, that “the greatest impact of open access is with the most-cited articles.”⁹

The desire to self-promote, especially those articles that one deems are worthy of additional exposure, may be a second explanatory cause in Antleman’s study, and explain why articles that she could find online were more highly cited than articles that were not. This explanation of causation works in the opposite direction – being online was the *result* of an author

promoting a high-impact article, or more explicitly, OA is the result of self-promotion.

Conclusion

The study of citation behavior is complex and involves multiple confounding, and interacting variables. Methodologically, it is very difficult to distinguish whether Open Access is an explanatory cause of increased access, or whether it is merely an artifact of other causal explanations such as article duplication or self-promotion. Do Open Access articles really have a greater research impact, as Antelman suggests? Yes, but Open Access may not be the cause. It may be more reasonable to say that “author republishing (online and in print) may increase citation impact, especially among highly prestigious journals and authors.” Although this is not as simple as declaring that Open Access increases citation impact, it may be much more precise.

Philip M. Davis
Life Sciences Librarian
Cornell University
pmd8@cornell.edu

Notes

1. Kristin Antelman, “Do Open Access Articles Have a Greater Research Impact?” *College & Research Libraries* 65, no. 5 (Sep 2004): 372-82.

2. *Ibid.*, 375

3. Philip M. Davis, “The Ethics of Republishing: A Case Study of Emerald/MCB University Press Journals,” *Library Resources & Technical Services* 49, no. 2 (2005): 72-78.

4. Philip M. Davis, “Article Duplication in Emerald/MCB Journals Is More Extensive Than First Reported: Possible Conflicts of Financial and Functional Interests Are Uncovered,” *Library Resources & Technical Services* 49, no. 4 (2005): 148-50.

5. Philip M. Davis, unpublished data.

6. Kent Anderson, John Sack, Lisa Krauss, and Lori Keefe, “Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience,” *Journal of Electronic Publishing* 6, no. 3 (2001), available: <http://www.press.umich.edu/jep/06-03/anderson.html>

7. Jonathan D. Wren, “Open access and openly accessible: a study of scientific publications shared via the internet,” *BMJ* 330 no. 7500 (2005), available: <http://bmj.bmjournals.com/cgi/content/abstract/330/7500/1128>

8. Wren. *Ibid.*, 4.

9. Antelman. *Ibid.*, 378

Response to Philip Davis

Philip Davis makes the important point that the data I presented in my paper are not sufficient to prove that open access causes citation advantage. While I intentionally phrased my conclusion as an association, rather than a causation (“open-access articles have a greater research impact than articles that are not freely available”), there clearly is an implied causation and I should have been more explicit that the data do not support that. The article was very much a product of its time, however, when there was little solid data that there even *was* an association between open access and increased citations.

At the same time the article was published, the topic of open access advantage causation began to be discussed more explicitly in papers and on e-mail lists. In September 2004 Stevan Harnad proposed a model that enumerated a number of components of open access advantage, including among others “early advantage,” “usage advantage,” and “quality bias.” Davis’s observed article duplication advantage likely has elements of both usage advantage and quality bias (Emerald may have tended to publish the better articles twice). In fact, article duplication

as a cause of increased citations is not incompatible with open access as a cause, since, after all, an open access copy is a duplicate copy.

The self-promotion cause (“quality bias”) does indeed, as Davis says, undermine any argument making a causal link between open access and citation advantage. Since I did the study in *C&RL*, I have collected additional data that indicate that quality bias is real and significant, at least in the social sciences. Wren’s study needs to be looked at carefully, however, because he did not look at the *source* of the open access copies he found, so the extent of “trophy effect” self-archiving cannot be assessed from his data. I also collected some data on the source of open access copies from three of the high-impact journals he looked at—*NEJM*, *Science* and *Nature*—and, while many articles from those journals are freely available online, many or most are not posted by the authors themselves, in particular, *NEJM* where only 12% of the open access articles were posted by authors or their institutions.

Kristin Antelman
Associate Director for Information
Technology
NCSU Libraries

Books & Periodicals ONLINE

The most comprehensive directory specifically developed for the database searcher seeking the online location of a source. Coverage includes both full text and bibliographic databases.

Also available in print in four topical editions:

- *Books & Periodicals ONLINE: Law, Business & News*
- *Directory of Periodicals Online: Science & Technology*
- *Directory of Periodicals Online: Medical & Pharmaceutical*
- *Directory of Periodicals Online: Humanities & Religion*

To set up a FREE trial, please contact us at info@booksandperiodicals.com or visit www.BooksandPeriodicals.com.

Library Technology Alliance, LTD
264 Lexington Avenue, New York, NY 10016
212.686.8816
212.686.8778 FAX