
Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial

Author(s): Philip M. Davis, Bruce V. Lewenstein, Daniel H. Simon, James G. Booth, Mathew J. L. Connolly and Godlee

Source: *BMJ: British Medical Journal*, Aug. 9, 2008, Vol. 337, No. 7665 (Aug. 9, 2008), pp. 343-345

Published by: BMJ

Stable URL: <http://www.jstor.com/stable/20510537>

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.com/stable/20510537?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



BMJ is collaborating with JSTOR to digitize, preserve and extend access to *BMJ: British Medical Journal*

JSTOR

- 11 Markowitz LE, Sepulveda J, Diaz-Ortega JL, Valdespino JL, Albrecht P, Zell ER, et al. Immunization of six-month-old infants with different doses of Edmonston-Zagreb and Schwarz vaccines. *N Engl J Med* 1990;322:580-7.
- 12 Garly ML, Martins CL, Balé C, da Costa F, Dias F, Whittle H, et al. Early two-dose measles vaccination schedule in Guinea-Bissau: good protection and coverage in infancy. *Int J Epidemiol* 1999;28:347-52.
- 13 Whittle H, Campbell H, Rahman S, Armstrong JRM. Antibody persistence in Gambian children after high-dose Edmonston-Zagreb measles vaccine. *Lancet* 1990;336:1048-51.
- 14 Samb B, Aaby P, Whittle H, Coll Seck AM, Rahman S, Bennett J, et al. Serological status and measles attack rates among vaccinated and unvaccinated children in rural Senegal. *Pediatr Inf Dis J* 1995;14:203-9.
- 15 Desgrées du Lou A, Pison G, Aaby P. The role of immunizations in the recent decline in childhood mortality and the changes in the female/male mortality ratio in rural Senegal. *Am J Epidemiol* 1995;142:643-52.
- 16 Leuridan E, Van Damme P. Passive transmission and persistence of naturally acquired or vaccine-induced maternal antibodies against measles in newborns. *Vaccine* 2007;25:6296-304.
- 17 Aaby P, Bukh J, Hoff G, Leerhøy J, Lisse IM, Mordhorst CH, et al. High measles mortality in infancy related to intensity of exposure. *J Pediatr* 1986;109:40-4.

Accepted: 19 May 2008

Open access publishing, article downloads, and citations: randomised controlled trial

Philip M Davis,¹ Bruce V Lewenstein,¹ Daniel H Simon,² James G Booth,³ Mathew J L Connolly⁴

EDITORIAL by Godlee

¹Department of Communication, Cornell University, Ithaca, NY 14853, USA

²Department of Applied Economics and Management, Cornell University

³Department of Biological Statistics and Computational Biology, Cornell University

⁴Cornell University Library

Correspondence to: P M Davis
pmd8@cornell.edu

Cite this as: *BMJ* 2008;337:a568
doi:10.1136/bmj.a568

ABSTRACT

Objective To measure the effect of free access to the scientific literature on article downloads and citations.

Design Randomised controlled trial.

Setting 11 journals published by the American Physiological Society.

Participants 1619 research articles and reviews.

Main outcome measures Article readership (measured as downloads of full text, PDFs, and abstracts) and number of unique visitors (internet protocol addresses). Citations to articles were gathered from the Institute for Scientific Information after one year.

Interventions Random assignment on online publication of articles published in 11 scientific journals to open access (treatment) or subscription access (control).
Results Articles assigned to open access were associated with 89% more full text downloads (95% confidence interval 76% to 103%), 42% more PDF downloads (32% to 52%), and 23% more unique visitors (16% to 30%), but 24% fewer abstract downloads (-29% to -19%) than subscription access articles in the first six months after publication. Open access articles were no more likely to be cited than subscription access articles in the first year after publication. Fifty nine per cent of open access articles (146 of 247) were cited nine to 12 months after publication compared with 63% (859 of 1372) of subscription access articles. Logistic and negative binomial regression analysis of article citation counts confirmed no citation advantage for open access articles.

Conclusions Open access publishing may reach more readers than subscription access publishing. No evidence was found of a citation advantage for open access articles in the first year after publication. The citation advantage from open access reported widely in the literature may be an artefact of other causes.

INTRODUCTION

In 2001 it was first reported that freely available online science proceedings garnered more than three times the average number of citations received by print

articles.¹ This “citation advantage” has since been validated in other disciplines. The primary explanation for this advantage is that freely available articles are cited more because they are read more than articles available by subscription only.

Some argue that open access articles are cited more because authors selectively choose articles to promote freely, or because highly cited authors disproportionately choose open access.²⁻⁵ This has been termed the self selection postulate.³ Self archiving an accepted manuscript in a subject based digital repository may provide additional time for these articles to be cited.³⁻⁵ A study of medical journals reported that the probability of an article being found on a non-publisher website was correlated with the impact factor of the journal.⁶ We carried out a randomised controlled experiment to measure the effect of free access to the scientific literature on article downloads and article citations from a journal publisher’s websites.

METHODS

From January to April 2007 we randomly assigned 247 research articles and reviews published in 11 journals of the American Physiological Society to open access status upon online publication (see bmj.com). The control group (1372 articles) was composed of subscription only articles.

We measured four different proxies for article readership: abstract downloads, full text downloads, PDF downloads, and unique internet protocol addresses. We also tested the effect of publisher defined open access on article citations (odds of being cited in year after publication and number of citations for each article).

Each month after publication we gathered usage statistics from the publisher’s websites. HighWire Press, the online host for the American Physiological Society’s journals, was able to provide reports on article downloads including and excluding internet robots. Article metadata (attributes of the article) and

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a568

citations were provided by the Institute for Scientific Information's Web of Science database. We carried out a search algorithm to identify as many instances of self archiving on the internet as possible.

Statistical analysis

We used linear regression to estimate the effect of open access on article downloads and unique visitors. Our outcome measures were downloads of abstracts, full text, and PDFs, and the number of unique internet protocol addresses (number of visitors). We log transformed these variables because of known skewness.⁷ Our principal explanatory variable was open access. We controlled for three influences on downloads: self archiving, being featured on a front cover, and being press released by the journal. We also controlled for article type, number of authors, authors based in the United States, number of references, length of article (pages), and journal impact factor. As articles are published within issues we nested the issue variable within the journal variable.

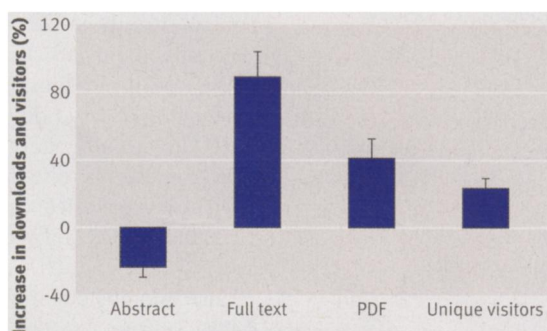
On 2 January 2008 we retrieved the number of article citations from the Web of Science. As our trial included articles published at different times (January to April 2007), we used a numerical indicator for each issue.

We estimated the effect of open access on citation counts using a negative binomial regression model with the same set of explanatory variables as described. This model is appropriate for count data and can work with over-dispersion in data.⁸

Finally, we used a logistic regression model to estimate the effect of open access on the odds of being cited, with the same set of explanatory variables employed in the download and negative binomial regression citation model.

RESULTS

The figure shows the effect of open access on article downloads and unique visitors in the first six months after publication. Full text downloads were 89%



Percentage differences (95% confidence intervals) in downloads of open access articles (n=247) and subscription access articles (n=1371) during the first six months after publication

higher (95% confidence interval 76% to 103%), PDF downloads 42% higher (32% to 52%), and unique visitors 23% higher (16% to 30%) for open access articles than for subscription access articles. Viewing of abstracts was 24% lower (-29% to -19%, $P < 0.001$ for all) for open access articles.

For open access articles, known internet robots could account for an additional 83% full text downloads, 5% additional PDF downloads, 4% additional unique visitors, and a 12% reduction of abstract downloads.

Regression analysis showed that several characteristics of articles had as much, or more, of an effect on article downloads as free access (see bmj.com). Having an article press released by the publisher increased PDF downloads by 65% (7% to 156%), and having an article featured on the front cover increased PDF downloads by 64% (21% to 121%). Longer articles, articles with more references, and those published in journals with higher impact factors had significantly more downloads.

Twenty instances of self archiving were identified—18 were final copies from the publisher and two were final manuscripts. The estimated effect of self archiving was positive on PDF downloads although non-significant (6%, -6% to 19%; $P = 0.36$) and essentially zero for full text downloads (-1%, -23% to 27%; $P = 0.95$).

Of the 247 articles randomly assigned to open access status, 59% (n=146) were cited after 9-12 months compared with 63% (859 of 1372) of subscription access articles.

The negative binomial regression model estimated that open access reduced expected citation counts by 5% (incident rate ratio 0.95, 95% confidence interval 0.81 to 1.10; $P = 0.484$), and that self archiving reduced expected counts by 10% (0.90, 0.53 to 1.55; $P = 0.716$, see bmj.com).

A supplementary logistic regression analysis based on the same set of variables estimated that open access publishing reduced the expected odds of being cited by about 13% (odds ratio 0.87, 95% confidence interval 0.66 to 1.17; $P = 0.36$, see <http://hdl.handle.net/1813/11049>), although this effect was not statistically significant.

DISCUSSION

Strong evidence suggests that open access increases the readership of articles but has no effect on the number of citations in the first year after publication. These findings were based on a randomised controlled trial of 11 journals published by the American Physiological Society.

Although we missed citation activity that occurred after these initial months, we believe that our time frame was sufficient to detect a citation advantage, if one exists.

Previous studies have relied on retrospective and uncontrolled methods to study the effects of open access. As a result they may have confused causes and

WHAT IS ALREADY KNOWN ON THIS TOPIC

Studies suggest that open access articles are cited more often than subscription access ones

These claims have not been validated in a randomised controlled trial

WHAT THIS STUDY ADDS

Open access articles received more downloads but exhibited no increase in citations in the year after publication

Open access publishing may reach more readers than subscription access publishing

The citation advantage of open access may be an artefact of other causes

effects or been unable to control for unmeasured variables.

Our finding that open access does not result in more citations suggests that the citation advantage may be an artefact of other explanations such as self selection.

To contribute meaningfully to the scientific literature access to resources as well as to the relevant literature are needed. These two requirements are highly associated and concentrated among the elite research institutions worldwide.^{9 10} That we observed an increase in readership and visitors to open access articles but no citation advantage suggests that the increase in readership is taking place outside the community of core authors.

The increase in full text downloads for open access articles in the first six months after publication (figure) suggests that the primary benefit to the non-subscriber community is in browsing, as opposed to printing or saving, which would have been indicated by a commensurate increase in PDF downloads. The fact that internet robots were responsible for much of the initial increase in full text downloads (83%) compared with PDF downloads (5%) implies that internet search engines are helping to direct non-subscribers to free journal content.

The discussion over access and its effects on citation behaviour assumes that articles are read before they are cited. Studies on the propagation of citation errors suggest that many citations are merely copied from other articles.¹¹⁻¹³ Given the common behaviour of citing from the abstract (normally available for free), the act of citation does not necessarily depend on access to the article. Secondly, the rhetorical dichotomy of “open” access compared with “closed” access does not recognise the degree of sharing that takes place among an informal network of authors, libraries, and readers.

Our citation counts are limited to those journals indexed by Web of Science. Because this database focuses on the core journals in a particular discipline,

we missed citations in articles published in peripheral journals.

We measured the number of unique internet protocol addresses as a proxy for the number of visitors to an article. We implied that the difference in number of visitors between open access and subscription based articles represents the size of the non-subscriber population. A more direct method of calculating access by non-subscribers would be to analyse the transaction logs of the publisher and to compare the internet protocol addresses from subscribing institutions with the total list of addresses. Because of confidentiality issues we did not have access to the transaction logs.

Open access articles on the American Physiological Society's journals website are indicated by an open green lock on the table of contents page. These may signal something about the quality of the article to potential readers and therefore created a positive bias on download counts.

Finally, we do not understand whether providing open access to articles had any effect on the behaviour of authors as they promoted their work. We are carrying out similar randomised experiments with other journals where authors and readers are unaware of the access status of articles.

We thank Bill Arms, Paul Ginsparg, Simeon Warner, and Suzanne Cohen at Cornell University for their critical feedback on our experiment.

Contributors: See bmj.com.

Funding: Grant from the Andrew W Mellon Foundation.

Competing interests: None declared.

Ethical approval: This study was approved by the institutional review board at Cornell University.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Lawrence S. Free online availability substantially increases a paper's impact. *Nature* 2001;411:521.
- 2 Davis PM, Fromerth MJ. Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics* 2007;71:203-15.
- 3 Kurtz MJ, Eichhorn G, Accomazzi A, Grant C, Demleitner M, Henneken E, et al. The effect of use and access on citations. *Inf Process Manage* 2005;41:1395-402.
- 4 Kurtz MJ, Henneken EA. Open access does not increase citations for research articles from the *Astrophysical Journal*. 2007. <http://arxiv.org/abs/0709.0896>.
- 5 Moed HF. The effect of 'open access' upon citation impact: an analysis of ArXiv's condensed matter section. *J Am Soc Inf Sci Technol* 2007;58:2047-54.
- 6 Wren JD. Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ* 2005;330:1097-8.
- 7 Seglen PO. The skewness of science. *J Am Soc Inf Sci* 1992;43:628-38.
- 8 Hilbe JM. *Negative binomial regression*. Cambridge: Cambridge University Press, 2007.
- 9 Cole JR, Cole S. *Social stratification in science*. Chicago: University of Chicago Press, 1973.
- 10 Price DJS. Collaboration in an invisible college. *Little science, big science and beyond*. New York: Columbia University Press, 1986:119-34.
- 11 Clark T. Copied citations give impact factors a boost. *Nature* 2003;423:373.
- 12 Simkin MV, Roychowdhury VP. Stochastic modeling of citation slips. *Scientometrics* 2005;62:367-84.
- 13 Simkin MV, Roychowdhury VP. A mathematical theory of citing. *J Am Soc Inf Sci Technol* 2007;58:1661-73.

Accepted: 18 May 2008