Open access archiving and article citations within health services and policy research

Devon Greyson, Steven Morgan, Gillian Hanley, and Desy Wahyuni

Abstract: Promoting uptake of research findings is an objective common to those who fund, produce, and publish health services and policy research. Open access (OA) is one method being employed to maximize impact. OA articles are online, free to access and use. This paper contributes to the growing body of research exploring the “OA advantage” by employing an article-level analysis comparing citation rates for articles drawn from the same, purposively selected journals. We used a two-stage analytic approach designed to test whether OA is associated with (1) the likelihood that an article is cited at all and (2) the total number of citations an article receives, conditional on being cited at least once. Adjusting for potential confounders (number of authors, time since publication, journal, and article subject), we found that OA archived articles were 60% more likely to be cited at least once and, once cited, were cited 29% more than non-OA articles.

Introduction

If health services and policy research is to make an impact on society, it must be accessible not only to academics, but also to policy-makers and the public. Promoting the uptake of research findings is an objective common to those who fund, produce, and publish health services and policy research. The prospect that all individuals with Internet access could have ready access to scholarly research publications is exciting and, at the same time, challenging. In an attempt to increase the diffusion of research outputs, the Canadian Institutes of Health Research is now among the many biomedical research funders that require grant recipients to “make every effort” to make research findings available on an open access (OA) basis [1].

OA articles are “digital, online, free of charge, and free of most copyright and licensing restrictions” [2]. This may be achieved by either (i) publishing in a journal that makes its content available without access fees (this is sometimes called “gold” OA; journals that make only a portion of their articles gold OA are commonly called “hybrid” journals) or (ii) publishing in a traditional subscription- or fee-based journal while also “archiving” a version of the article elsewhere for people to access without paying a fee (“green” OA). Because OA is a new possibility (owing to the relative ease and affordability of disseminating information over the Internet), research publishers, producers, and funders are still grappling to understand its pros and cons.

There are ethical motivations for providing research findings on an OA basis. For example, it might be argued that taxpayers have a right to freely access the products of the institutions, researchers, and projects that receive public funds [3]. OA may also be motivated on the basis of expediency: to the extent that research aims to advance knowledge, foster innovation, and change policy and practice in any given domain, OA may accelerate those processes by dramatically reducing the financial and time costs associated with accessing knowledge [4]. Universal access to research findings, therefore, may not only democratize but also advance discourse about a particular subject.

There is also the argument that providing OA to articles, thereby removing both cost and time barriers to accessing the literature, increases the impact of the research. Some refer to this notion—that OA increases the impact of research—as the OA advantage or OA impact advantage [5,6]. Although it may take years to establish whether OA ultimately accelerates innovation and changes in practice, process measures related to the OA advantage can be investigated via empirical methods. This has been studied in a growing number of scholarly fields, most notably those that were early adopters of electronic publishing and archiving such as astronomy [7] and computer science [8].


S. Morgan. UBC School of Population and Public Health; and UBC Centre for Health Services and Policy Research, Vancouver, BC V6T 1Z3, Canada.

D. Wahyuni. Courthouse Libraries BC, 800 Smithe Street, Vancouver, BC V6Z 2E1, Canada.

1This article has been peer-reviewed.

2Corresponding author (e-mail: devon@chspr.ubc.ca).
It is difficult, however, to generalize findings among disciplines, as communication norms may vary greatly. Both online technical capacity and the terminology to describe it have evolved quickly over recent years, and the very findability of relevant research has thus suffered by the lack of a controlled vocabulary in which to discuss the issue. Furthermore, measures of research impact are far from universally agreed upon.

At this time, there is little guidance available regarding what OA means for publication impact in the health services and policy field in Canada and beyond. Many Canadian publishers are relatively small in size, and several of our journals are not indexed in many major citation databases. Much Canadian health services and policy research is therefore published in large international journals. The health services and policy research field, like most interdisciplinary subject areas, has not yet been empirically investigated in the growing body of studies examining whether OA publishing or archiving has an effect on research impact.

**Previous studies**

There is a young but growing body of empirical research on the potential impact advantage of OA publishing. However, this literature is the subject of vigorous methodological debates, and there is not yet a consensus on any single “gold standard” method or article. Major challenges in this research include assessing the impact of a publication, determining appropriate comparators to the OA publications (for a given article is either OA or not OA), and untangling the complex web of factors that may contribute to the use or impact of a work. Disagreement is ongoing over such topics as whether studies should focus on articles or journals, whether it is appropriate to focus on gold or green OA (or both, and if so, how), and whether or not retrospective studies can shed useful light on the issue.

Although an imperfect measure of research uptake, the most commonly used test for an OA advantage is whether OA is correlated with the impact factor (IF) and/or article citation counts from the Institute for Scientific Information (ISI) Journal Citation Reports (JCR). The IF, a formulaic number that remains significant in many tenure review processes, has been much criticized for its biases, errors, and vulnerability to manipulation by editors [9,10], but it remains a longstanding measure of journal prestige still valued (if sometimes inappropriately applied) by many institutions [11]. Various alternative measures of journal and article impact have been developed and are being tested [12–14]. Alternative measures of use or influence may add another dimension or more accurately reflect the broader impact of an article outside the American-slanted set of articles in a specific discipline that are included in the JCR; however, none of these have clearly risen to the top as a new standard at this point.

Some studies have compared publication impact at the journal level [15,16]. These retrospective, observational studies suggest a mixed association between OA status and impact and are vulnerable to criticisms concerning whether it is possible or appropriate to match OA and non-OA publications on a journal-to-journal basis. No two journals are exactly comparable, and some differences, such as primary audience and editorial style, may be difficult for researchers to objectively measure and account for when attempting to measure any OA advantage. Moreover, as Harnad and Brody point out, there may be an “element of circularity” in matching journals according to all available criteria and then assessing whether those journals are quite similar in impact [17].

Other investigators have compared OA and non-OA articles published within the same (hybrid OA) journal. A relatively early study of OA found that articles that the editors of *Pediatrics* decided to make freely available as part of an online-only section of the journal in 1997–1999 were far more likely to be accessed online, though not more likely to be cited in the literature, than subscription-based articles from the same journal [18]. A more recent study indicated that articles in the *Proceedings of the National Academy of Sciences* (PNAS) for which authors had paid a fee to make available on an OA basis had a higher citation rate than print-only articles in the same journal [19], but the study found no citation advantage among green OA articles from the same publication that authors had archived elsewhere on the Internet. In both of the preceding studies, it is unknown whether articles that the editors or authors decided to post online differed in substantive ways from print-only articles in the same journals. Additionally, focusing on a particular journal necessarily limits the extent to which findings can be generalized.

A few studies have examined OA impact using panel data from large samples of articles from journals across subject groups [8,20–22]. All of these studies have found that OA articles have higher mean citation rates than non-OA articles within disciplines. One study also found that the difference in citation rates between OA and non-OA articles was greatest in mathematics—the discipline with greatest level of OA adoption—and lowest in philosophy, which had the least OA saturation [20]. This result suggests that as OA adoption within a field increases, non-OA articles may be at greater citation disadvantage when compared with those that are OA.

As is to be expected in the investigation of a relatively new phenomenon, most of the OA impact studies to date have been observational, and further, most of these have been retrospective (with the PNAS study [19] being a notable prospective exception). Looking at cross-sections of articles at a given time or observing articles within specified publications over time are useful ways to identify trends that may be significant. Although there is not complete consensus among these observational studies, experimental studies are beginning to test the resultant theories. Davis’s preliminary report of a randomized controlled trial of gold OA publishing [23] is the first publication reporting on such an experiment, but the report is inconclusive. It seems likely that within the next few years we will start seeing more experimental studies of OA publishing and (or) archiving that will shed further light on the extent to which OA is or is not a causative factor to various types of research impact.

The purpose of this study is to measure the potential OA advantage in the specific context of Canadian health services and policy research by assessing whether OA status increases the likelihood of an article being cited, as well as the association between OA status and the total number of times an article is cited.
Methods

Data
We employed a cross-sectional, retrospective article-level analysis comparing citation rates for articles drawn from the same, purposefully selected journals. Because our intention was to investigate the use and potential impacts of publishing options and policies used by Canadian health services and policy researchers, we selected journals of high use as publishing outlets for researchers in the field. To identify these journals, we surveyed the directors of nine health services and policy research centres across Canada. We asked these directors to list the 10 journals in which their researchers most frequently publish. Four directors provided responses to the survey (44% response rate). From the 24 unique journals provided by those responses, we selected all traditional access journals that were indexed in JCR and permitted author self-archiving of refereed articles. This gave us a list of four journals for investigation, all of which are international journals rather than publications with Canadian-specific focus: *Health Economics* (Wiley, 2005 IF = 1.919); *Health Policy* (Elsevier, 2005 IF = 0.964); *Journal of Health Politics, Policy and Law* (Duke University Press, 2005 IF = 0.718); and *Social Science & Medicine* (Elsevier, 2005 IF = 2.619).

For each of these four journals, we exported citation reports for all original articles published from 2003 to 2005 from ISI’s Web of Science. These reports included bibliographic data on each article as well as a record of citations to the article in each subsequent year. As we obtained these records in July of 2007, we are reasonably confident that we have complete citation data for the 3 years following publication of articles published in 2003, 2 years following articles published in 2004, and 1 year for those published in 2005. These lists were validated against PubMed’s MEDLINE database to catch any “dropped” articles, resulting in a final set of 1923 articles.

Within the selected journals, we tested for an OA impact for articles that had been archived by authors versus other articles in the same journal. Using the article titles, we searched in Google (including Google Scholar) and PubMed in July 2007 to locate any OA archived copies of the articles. Articles for which archived copies could be found online (live or cached) were classified as OA; all other articles from the four journals were classified as non-OA. Our analysis therefore measures the impact of green, not gold, OA; anyone interested in reading the articles in our study would still have to find the archived copies or pay traditional access fees. Further, as this was a retrospective study and not a longitudinal study over time, this variable is an “ever OA” marker, rather than an indicator of how long an article had been archived.

We classified articles into subject areas within health services and policy research by using the EBSCO interface to MEDLINE to retrieve subject headings deemed to be major topics of each article in our set (MeSH MAJR). These major subject headings were collapsed into 13 broad categories using the National Library of Medicine’s subject taxonomy (MeSH Tree Structures 2007). Each article in our study was mapped to one or more of the broad subject areas. Subject areas pertaining to fewer than 50 of the articles in our sample were dropped from the regression analyses.

Statistical methods
We used a two-stage analytic approach designed to test whether OA is associated with (1) the likelihood that an article is cited at all and (2) the total number citations that an article receives, conditional on the article being cited at least

### Table 1. Summary of articles included in this study by journal and subject category.

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of articles</th>
<th>% available OA*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Journal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Science &amp; Medicine</td>
<td>1220</td>
<td>16.1</td>
</tr>
<tr>
<td>Health Economics</td>
<td>277</td>
<td>29.2</td>
</tr>
<tr>
<td>Health Policy</td>
<td>337</td>
<td>16.6</td>
</tr>
<tr>
<td>Journal of Health Politics, Policy and Law</td>
<td>89</td>
<td>25.8</td>
</tr>
<tr>
<td><strong>Subject area</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychiatry/psychology</td>
<td>1026</td>
<td>15.9</td>
</tr>
<tr>
<td>Diseases</td>
<td>480</td>
<td>16.6</td>
</tr>
<tr>
<td>Chemicals/drugs</td>
<td>89</td>
<td>18.0</td>
</tr>
<tr>
<td>Analytical/diagnostic/therapeutic techniques and equipment</td>
<td>672</td>
<td>18.9</td>
</tr>
<tr>
<td>Biological sciences</td>
<td>862</td>
<td>19.5</td>
</tr>
<tr>
<td>Anthropology/education/sociology/social phenomena</td>
<td>1004</td>
<td>19.6</td>
</tr>
<tr>
<td>Information science</td>
<td>266</td>
<td>16.5</td>
</tr>
<tr>
<td>Named groups of persons</td>
<td>425</td>
<td>18.4</td>
</tr>
<tr>
<td>Health care: population characteristics</td>
<td>606</td>
<td>20.6</td>
</tr>
<tr>
<td>Health care: facilities/manpower/services</td>
<td>574</td>
<td>15.2</td>
</tr>
<tr>
<td>Health care: economics and organization</td>
<td>749</td>
<td>21.8</td>
</tr>
<tr>
<td>Health care: health services administration</td>
<td>675</td>
<td>18.0</td>
</tr>
<tr>
<td>Health care: quality/access/evaluation</td>
<td>1006</td>
<td>18.8</td>
</tr>
</tbody>
</table>

*OA, open access, which is defined here as author self-archiving of published material (i.e., green OA).
once. In the first stage of our analysis, we used a logistic regression as a measure of the association between OA status and the likelihood that the article has been cited. We then ran a second stage model of the total number of citations, conditional on an article being cited once.

Both stages of our model adjusted for factors that might affect citation rates, irrespective of OA status: the number of authors on the article, the time since publication, the journal, the subject heading categories, and the interactions between time and journal (dummy variables). The log-likelihood ratio test was used to assess the significance of the association fitted by the logistic regression model. The $F$-test was used to assess the significance of the two-stage model. Statistical analyses were performed in Stata version 10.0 (StataCorp LP, College Station, Tex.).

To assess whether differences in citation rates between OA and non-OA articles can be explained by open access articles receiving citations earlier than non-OA articles, we examined the subsample of articles published in 2003. We examined the yearly citation rate for OA and non-OA articles for the years 2003, 2004, 2005, and 2006 and used the $t$-test to determine whether the differences between the citation rates were significantly different in each year.

**Results**

Table 1 summarizes our study sample, providing the total numbers of articles studied and the percentage available on OA terms by journal and subject area. We collected information on all 1923 “citable” articles published in all four journals from the year 2003 to 2005, including citations to the articles as indexed in Web of Science through July 2007. We found OA archived versions of 357 (18.6%) of the articles in this study. The rate of such OA availability was higher among articles in *Health Economics* (29.2%) and the *Journal of Health Politics, Policy and Law* (25.8%) than in *Social Science & Medicine* (16.1%) and *Health Policy* (16.6%). Differences in OA availability across subject areas were more modest, ranging from 15.2% (health care: facilities/manpower/services) to 21.8% (health care: economics and organization).
Overall the proportion of articles that were cited at least once was 85%. The overall frequency at which articles were cited differed by OA status. Approximately 90% of OA articles were cited at least once, whereas 84% of the non-OA articles were cited one or more times. Without adjusting for journal, subject, or time, the OA status increases the probability of ever being cited by 60% (crude odds ratio (OR) = 1.60; 95% confidence interval (CI): 1.11, 2.30).

Table 1 and Table 2 illustrate the archiving practices used among the OA portion of our sample. Most OA archived articles were archived in only one location, while 27.5% were found in multiple locations online. The majority of these archived copies (75.9%) were found on personal or institutional Web sites, with only 17.0% found in institutional or subject-based OA repositories.

Multivariate logistic regression

Adjusting for journal, subject, and time of publication did not significantly alter the principal finding. Table 3 presents the odds ratios from the multivariate logistic regression of the likelihood that articles are cited at least once. The log-likelihood ratio (LLR) of this logistic regression model indicates that at least one coefficient is significantly different from zero (LLR = 200.22 with 20 degrees of freedom; p < 0.0001). The results suggest that, after adjusting for factors related to the journal, subject matter, and article, OA status is associated with a statistically significant 64% increase in the likelihood that an article will be cited at least once (OR = 1.64; 95% CI: 1.11, 2.45).

Articles appearing in *Health Policy* were less likely to be cited at least once than articles in *Social Science & Medicine* (the comparator). Articles in *Health Economics* and the *Journal of Health Politics, Policy and Law* were no more or less likely to be cited than those in *Social Science & Medicine*. In comparison with the subject heading of “psychiatry/psychology”, there were few statistically significant subject-level differences: articles about “chemicals/drugs” were more likely to be cited at least once, while articles about “named groups of persons” were slightly less likely to be cited.

Given the theory that the relationship between OA status and likelihood of citation and number of citations is confounded by the number of authors, we ran a separate regression analysis examining the relationship between the number of authors and OA status. Both unadjusted and adjusted analyses suggest that there is no significant increase in the likelihood of being OA status as number of authors increase. However, we did find that after adjusting for time, journal, and article subjects, articles with more authors are more likely to be cited one or more times than articles with fewer authors.

Table 4 lists the results of the two-stage model of association between an article’s OA status and its total number of citations, conditional on the article having at least one citation. We find a statistically significant positive association between OA status and the total number of citations. Controlling for journal, number of authors (one, two, three, or four or more authors), time since publication, and subject,
cited OA articles have 1.67 more citations than cited non-OA articles. Given that the average number of citations among all articles in our study was 5.71, these results would suggest that open access status is associated with a 29% increase in number of citations.

In comparison with cited articles in Social Science & Medicine, which is both our reference journal and the journal with the highest impact factor within our sample, cited articles in the other three journals included in this study had lower total citations. In comparison with the subject heading of “psychiatry/psychology”, the only statistically significant finding was that articles indexed as “named groups of persons” or “health care: health services administration” had fewer total citations.

Figure 2 outlines the average yearly citation rates in our subsample of 2003 articles (589 articles). OA articles have a higher average yearly citation rate both immediately (in 2003) and consistently through the study period (2003–2006). Our statistical tests (t-test) indicate that OA articles published in 2003 have a statistically significantly higher average rate of citations than non-OA articles in the years 2004, 2005, and 2006, and that difference in average citation rate between OA and non-OA articles increases over time. Trends in citations to articles published in 2004 and 2005 appear to be similar, but the 3-year follow-up period for those publication years was not yet complete at time of analysis.

Discussion

Our study objectives were to determine whether OA articles were more likely to be cited, and, given that they were ever cited, whether they received a higher total number of citations than non-OA articles. Our investigation used an appropriate comparison group (articles within the same journal as opposed to comparing articles across journals, articles without a difference in publication fees as opposed to comparing paid-for articles with no-fee articles) and controlled for a potential confounder—individual article subjects within a discipline, using MeSH—that had not been controlled for in previous studies.

The findings presented above suggest that among key health services and policy research journals, articles that are made available on OA terms through green archiving are more likely to be cited and, once cited, are likely to receive a larger number of citations than similar non-OA papers. Articles in higher impact journals within our sample were naturally more likely to be cited and received more total citations, and each additional author of an article has a significant and positive impact on total citations. Our results are consistent with prior literature examining citation patterns [24–26] and the effect of open access status [15,20,21].

Limitations

Our study sample and the methods we employed necessarily place some limitations on our conclusions. The retrospective, cross-sectional design allowed us to get a “snapshot” of the state of the literature in our sample, and to raise questions about, but not draw conclusions regarding, the causes of any impact differential we observed. This study focused exclusively on green OA, archived copies of traditionally published articles; therefore, our conclusions may not be applicable to other forms of OA such as gold OA publishing. The study’s small sample size of 1923 articles limits our statistical power. One area in which this is quite evident is in the analysis of article subject, in which we had to collapse the more than 1400 subject headings into just 13 broad categories. A larger study might be able to identify trends within small research communities that were not observable in our investigation. Another limitation of the study is the relatively short time frame. Although 3 years is all the time required to see the effect of citations upon a journal’s impact factor, we do not yet have enough data to rigorously assess any impact of OA status upon long-term citation trends. Thus, while we know that most citations to an article are usually made soon after publication, we do not know whether OA might provide an additional advantage to articles by way of increased persistence of citations (a “longer tail”) into the more distant future.

We focused on ISI’s citation reports and the impact factor as measures of impact, although other measures of impact and use are increasingly available [14,9,27]. It is undeniable that in a field like health services and policy research, which generally intends for its research outputs to reach beyond academic circles and across disciplines, counts of citations made by other scholarly articles within the discipline is an incomplete measure of impact. It is further recognized that the ISI indices are US biased and thus do not count some portion of the Canadian scholarly impact of the articles in this study. However, taking into account the criticisms of the ISI citation indices and impact factor, we chose to use it in our investigation as it remains the largest, most-established citation index of scholarly literature available, and the journal impact factors remain significant to publishers and scholars alike.

Additionally, there are some desirable data that would have enriched the study were we able to access and use such information to develop controls for our analysis. One of these is the time elapsed since archiving. Our ever-archived versus never-archived OA variable is unable to capture nuances based in citations to an article over time since archiv-
ing. More author information would also have added to our analysis (e.g., measures of author prestige, author affiliations, author funding, and author’s stated reasons for archiving or not archiving). Such author-specific information would help investigate the theory that authors may selectively make their best work OA and (or) that other factors such as study funder or author institution may cause both OA and higher citations to an article, independently of each other.

Conclusion

This study’s intent was to examine citations to articles within health services and policy research journals of high interest to Canadians to assess the possibility and potential magnitude of any open access (OA) advantage based on self-archiving of peer-reviewed, published research articles. While the ultimate measure of the value of OA publication is whether it enhances the pace and extent to which research affects policy and practice, our study shows that OA is associated with process measures (scholarly citations) indicating greater diffusion of research results. We found that in journals of high interest to Canadian health services and policy researchers, OA archiving of peer-reviewed research articles is associated with a 64% greater likelihood of being cited as well as a 29% higher citation count among the cited articles.

There thus appears to be a citation advantage associated with OA archiving of health services and policy research articles. Whether this is caused by the articles’ OA status, confounding factors, or some measure of both, however, remains to be teased apart by further studies. Recommendations for future research to more fully investigate the impact of OA within health services and policy research include the following: (i) studies of the information behaviour of non-academic users of this body of research (e.g., health policy decision makers, the media, and the general public); (ii) further studies of authors (author motivations for archiving articles, author archiving and OA publishing behaviour patterns, and factors influencing those behaviours); and (iii) prospective experimental studies of the impact of OA, applied randomly to articles in matched cohorts (e.g., randomized controlled trials of articles within journals).

Despite the limitations of the current study, we believe recommendations can be given to researchers, publishers, and funders of health services and policy research based on our findings, placed within the context of those in the broader literature on OA publication. Although we cannot claim to know for certain that green OA causes the higher citations with which it is associated, self-archiving is typically free to do and requires minimal time and effort on the part of the author or author’s liaison. The correlation of OA archiving with higher citations, when balanced with both the low investment required to self-archive and the ethical argument for OA, makes a compelling argument for providing open access to our health services and policy research literature.

References

4. Swan A. Open access and the progress of science: the power to transform research communication may be at each scientist’s fingertips. Am Sci. 2007;95.


22. Citation advantage for OA self-archiving is independent of journal impact factor, article age, and number of co-authors—open access archivangelism. 2007 [cited 2007 Mar 26]. Available from: http://eprints.ecs.soton.ac.uk/13329/01/eysen.html.


