Do open access journal articles experience a citation advantage? Results and methodological reflections of an application of multiple measures to an analysis by WoS subject areas

Isabel Basson[1] · Jaco P. Blanckenberg[1] · Heidi Prozesky[1]

Abstract

This study is one of the first that uses the recently introduced open access (OA) labels in the Web of Science (WoS) metadata to investigate whether OA articles published in Directory of Open Access Journals (DOAJ) listed journals experience a citation advantage in comparison to subscription journal articles, specifically those of which no self-archived versions are available. Bibliometric data on all articles and reviews indexed in WoS, and published from 2013 to 2015, were analysed. In addition to normalised citation score (NCS), we used two additional measures of citation advantage: whether an article was cited at all; and whether an article is among the most frequently cited percentile of articles within its respective subject area (pptopX %). For each WoS subject area, the strength of the relationship between access status (whether an article was published in an OA journal) and each of these three measures was calculated. We found that OA journal articles experience a citation advantage in very few subject areas and, in most of these subject areas, the citation advantage was found on only a single measure of citation advantage, namely whether the article was cited at all. Our results lead us to conclude that access status accounts for little of the variability in the number of citations an article accumulates. The methodology and the calculations that were used in this study are described in detail and we believe that the lessons we learnt, and the recommendations we make, will be of much use to future researchers interested in using the WoS OA labels, and to the field of citation advantage in general.

Keywords Citation analysis · Open access citation advantage · Open access journals · Web of Science · Normalised citation scores · Open access

✉ Isabel Basson
     ibasson@sun.ac.za

1   DST-NRF Centre of Excellence in Scientometrics and STI Policy; and Centre for Research on Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa

Introduction

The Budapest Open Access Initiative (BOAI) declaration (Chan et al. 2002) contains one of the first and most definitive definitions of open access (OA) to peer-reviewed journal publications, and it lay the foundation of the OA movement (Miguel et al. 2016, p. 7). Sentiments in this declaration that proclaim the benefits of OA for authors of scientific literature and for their institutions and readers, lie at the core of the calls made by OA proponents for support of OA. Some of the suggested benefits of OA include accelerated research, enriched education, reduced costs to disseminate research, and increased visibility and impact of research. With an increase in prevalence of OA peer-reviewed journal publications since the declaration, and a new method to identify OA journal articles, it is now possible to investigate, on a sufficiently large scale, whether the proclaimed benefits have indeed been realised. We investigate the veracity of the claim of the BOAI declaration that OA provides a "*vast and measurable*" increase in "*impact*" and "*visibility*" of peer-reviewed journal publications, as measured through citations (Chan et al. 2002, p. 1), i.e. that those publications experience an OA citation advantage (Kurtz et al. 2005, p. 11). This claim is referred to as the OA citation advantage postulate (Craig et al. 2007). We specifically investigate the OA citation advantage postulate as it pertains to OA journal articles, in comparison to subscription journal articles. In order to control for potential differences in citation behaviour across OA types, we limited our results analysis to only one type of OA, excluding other types of OA, e.g. OA self-archived articles and hybrid OA articles.

Our investigation involved analysing bibliometric data on articles published from 2013 to 2015, as indexed in the Clarivate Analytics Web of Science (WoS) citation index, and disaggregating by WoS subject area. It is one of the first studies that use the recently introduced OA labels in the metadata and as such includes a description of the characteristics of the OA label as used by WoS. Using those labels allowed us to identify OA journal articles, and to distinguish between subscription journal articles for which self-archived versions are available, and those for which no such versions are available. In addition, these labels allowed us to investigate the previously anecdotal observation that OA journal articles are more likely than subscription journal articles to be written in a language other than English. The study improves on previous studies on the OA citation advantage by measuring not only the mean number of citations articles have received, but by using two additional measures of citation advantage, viz whether an article was cited at all; and whether an article is among the most frequently cited percentile of articles within its respective subject area (pptopX %). As far as we could ascertain, these two measures have rarely been applied to the investigation of the OA citation advantage.

A conceptual and empirical review of the open access citation advantage

The notion that OA provides a measurable increase in visibility is based on the idea that OA articles are available to a larger audience than they would be if published in non-OA outlets, as access is not limited by a "*paywall*" and as such, the audience of OA articles includes anyone that has access to the internet (Harnad et al. 2008). Visibility, readership and impact are often measured by citations and, arguably, researchers can only use and cite articles that they can read. Thus, the basis of the OA citation advantage postulate is that

articles that are rendered OA will receive more citations than articles of the same quality, but to which access is limited, and that this is due to the increase in the size of the audience able to cite the OA articles (Cullen and Chawner 2011, p. 463). It is important to note that, in both cases, the articles still need to report on original, quality research to attract citations (Gargouri et al. 2010, p. e13636).

Since the early 2000s, various studies have been conducted to examine whether OA leads to a citation advantage (e.g. Lawrence 2001; Antelman 2004; Kurtz et al. 2005; Hajjem et al. 2005, Sotudeh and Horri 2007). These initial studies not only assisted with identifying the mechanisms through which an OA citation advantage could possibly be explained (Davis and Fromerth 2007; Sotudeh and Horri 2009), but also found that an OA citation advantage differs across the various types of OA (e.g. green, gold, hybrid) (Archambault et al. 2014b, p. 20; Periances-Rodríguez and Olmeda-Gómez 2019, p. 1743) and subject areas (Dorta-González et al. 2017). However, there is still no consensus whether an OA citation advantage exists, as previous studies, and therefore also their results, differ in terms of the types of OA and subject areas investigated, as well as the sampling methods, and citation indexes used.

The introduction of additional data labels associated with OA into WoS facilitated new research into OA publications. In 2014, WoS introduced a data label which allowed for the identification of articles published in OA journals listed on the Directory of Open Access Journals (DOAJ) (Torres-Salinas et al. 2016). In 2017, Unpaywall data were added to the WoS citation index, which expanded the classification of OA articles to include additional types of OA (particularly green OA) and classified articles individually and no longer simply based on the journal it is published in (Basson 2019). This opened up new possibilities to investigate OA and the OA citation advantage. These labels were again updated in 2019, as the Unpaywall data underwent additional changes resulting in additional OA labels introduced in the WoS metadata (Clarivate Analytics 2018; Basson 2019). We made use of these new opportunities to investigate the whole population of articles indexed in WoS (published from 2013 to 2015), while many of the previous studies referred to above, had to rely on smaller samples. We also improve on the previous studies by using multiple measures to investigate the OA citation advantage postulate, which has previously been measured primarily by comparing OA articles to non-OA articles in terms of the mean number of citations they receive (e.g. Dorta-González et al. 2017; Van Leeuwen et al. 2018). As the distribution of citations across articles is severely skewed, the mean number of citations of any grouping of articles is influenced by a small group of extremely frequently cited articles (Porter 1977, p. 263). It has therefore been recommended that other measures of citation impact should also be used when conducting citation analysis using citation counting (Waltman 2016, p. 371).

We used two alternative measures that have been suggested in the literature, namely an article's presence among (1) uncited articles; and (2) frequently cited articles. With regard to the former, it is relevant to note that receiving any citations is often seen as a measure of visibility (Cronin 1984, p. 25), while articles that are never cited are sometimes viewed as obsolete and as not contributing to science (Burrell 2002, p. 232; Cronin 1984, p. 32). As such, determining whether OA articles are less likely to remain uncited than their non-OA counterparts can be considered a valid measure of the OA citation advantage. It complements other measures of citation advantage, as it could be argued that, while a wider audience does not necessarily lead to an article receiving an exceptionally high number of citations, the increased visibility could lead to an article being more likely to be cited at all. As far as we can ascertain, this measure of citation advantage has not been used previously to directly investigate the OA citation advantage postulate. However, a few studies have

investigated uncited articles within the context of OA, and those found that OA journal articles in some subject areas are less likely to remain uncited than their non-OA counterparts. For example, a study of WoS-indexed articles published from 2001 to 2003 in the subject areas 'Engineering and material sciences', 'Life sciences', 'Multidisciplinary sciences', and 'Natural sciences', found that a smaller percentage of articles published in OA journals remained uncited in comparison to articles published in non-OA journals with the exception of those in 'Life sciences' (Sotudeh and Horri 2007, p. 2148). Similar results have been reported for hybrid OA articles published in 2004 in the Proceedings of the National Academy of Sciences of the United States of America (PNAS), using citation data from WoS (Eysenbach 2006, p. 693).

The second alternative measure that we used, is a percentile-based indicator of the presence of articles among frequently cited articles. Such indicators are commonly used to investigate the most frequently cited articles (Waltman and Schreiber 2013, p. 372; Tijssen et al. 2002; Van Leeuwen et al. 2003). It is therefore quite surprising that this indicator has rarely been used when investigating whether OA journal articles experience a citation advantage. The few such studies that we identified mostly investigated types of OA other than OA journal articles (e.g. Gargouri et al. 2010 studied self-archived articles).

However, three exceptions merit discussion here. First, Torres-Salinas et al. (2016) investigated the percentage of OA journal and non-OA journal articles among the 10% most frequently cited articles published in selected disciplines by authors with a Spanish affiliation, using WoS for citation data and WoS's initial OA labels to identify OA journal articles. They found that OA journal articles did not experience a citation advantage. Secondly, Fukuzawa (2017) investigated the 10% most frequently cited articles published from 2010 to 2012 in OA journals and non-OA journals, using Scopus for citation data and the DOAJ and Open Access Scholarly Resources (ROAD) to identify OA journal articles. The study found a citation advantage, but for non-OA journal articles, and only if they were published in international journals. The study also suggests that factors other than whether an article is published in an OA journal might have a larger effect on whether an article is highly cited or not, such as whether the article is published in English. This provided further justification for applying a percentile-based indicator of the presence of articles among frequently cited articles and to investigate the language an article is published in as a potential confounding variable when investigating the OA citation advantage postulate. Thirdly, and more recently, Perianes-Rodríguez and Olmeda-Gómez (2019) investigated whether the proportion of articles among the 1% most frequently cited articles differs between OA journal articles and other articles. They used data from Scopus, WoS, SHERPA/RoMEO, and the DOAJ for selected subject areas. The study found no evidence of a citation advantage for OA journal articles.

In summary, we identified very few studies on the OA citation advantage postulate that made use of measures other than mean number of citations, especially when investigating OA journal articles. Those studies that did, each made use of different data sources and methods to identify OA articles and non-OA articles. In addition, they either investigated selected subject areas (disaggregating amongst those) or included all subject areas without disaggregation on this variable. Disaggregation by subject area is, however, crucial, as citation behaviour differs significantly among subject areas (Waltman et al. 2011, p. 467). Our review of the literature highlights the need to include the investigation of language as a confounding variable; use measures other than mean number of citations; and consider subject areas individually when investigating the OA citation advantage of OA journal articles. Previously it was impractical to investigate measures of citation advantage for all subject areas, due to the difficulties in identifying OA journal articles and self-archived articles

across a large set of articles. Using the updated Unpaywall OA labels recently added to the WoS metadata, this study aims to overcome these limitations and contribute to a more comprehensive understanding of the citation behaviour of OA journal articles.

Methodology

Research questions

The central question this study aimed to address is whether OA journal articles within a particular subject area experience a citation advantage, in comparison to subscription journal articles in the same subject area. As explained above, we apply three different measures of citation advantage, which translates into the following three sub-questions:

- Are the normalised citation scores (NCSs) of OA journal articles higher than those of subscription journal articles?
- Is the percentage of OA journal articles that receive at least one citation within 2 years after publication higher than that of subscription journal articles?
- Are the percentages of OA journal articles among the most frequently cited 1%, 5%, and 10% of articles higher than those of subscription journal articles?

In addition, we investigated whether there is a relationship between whether an article is published in English and whether it is published in an OA journal. The result of this analysis informed our decision to include or exclude English-language publications from the rest of the analysis, depending whether language was identified as a confounding variable.

In this "Methodology" section, we begin by describing our operationalisation of the three measures implied in these research questions. This is followed by a sub-section on our conceptualisation and operationalisation of 'access status'. The methodology section concludes with a discussion of our data extraction methods in a sub-section that includes our decisions regarding the definition of the population and variables relevant to that definition.

Measures of citation advantage

This study improves on previous research on the OA citation advantage of OA journal articles in comparison to subscription journal articles, by applying three measures of citation advantage, two of which have been rarely used. In the following three sub-sections, we provide more detail on each of these measures.

Normalised citation score

The first measure of citation advantage we used is based on the NCS (Waltman et al. 2011, pp. 469–470), which 'normalises' or corrects for both subject area and publication year. An article with an NCS of less than one (the average for its subject area and publication year) indicates that the article received below the expected number of citations, while an NCS of more than one indicates the opposite. Since articles are often associated with more than one subject area, each article and all citations it receives is attributed in equal fractions to all the subject areas associated with it. We selected a 6-year citation window, which

allowed the counting of citations to articles in subject areas in which it takes longer to accumulate citations, while still normalising by publication year to compensate for variation in the number of years since publication that articles could potentially be cited. While 6-years have not passed for all of the articles, the normalisation results in articles still being comparable. Self-citations were included in the calculation of the NCS, as the complicated process of accurately identifying and removing self-citations was not deemed justified, considering evidence that excluding them has little effect on citation analysis conducted on highly aggregated levels, such as the analysis reported in this article (Aksnes 2003, p. 244; Glänzel and Thijs 2004, p. 286).

In order to test the OA citation advantage postulate, we determined whether the NCSs of OA journal articles are higher than those of subscription journal articles, by consulting the effect size of the relationship between the NCS and the variable access status. For this continuous-by-dichotomous relationship, the point-biserial correlation coefficient (r_{pb}) was used, and a result of $|r_{pb}| \geq 0.1$ was interpreted as a large enough effect size to indicate a relationship, $|r_{pb}| \geq 0.3$ is taken to indicate a strong relationship. A negative value indicates an OA citation advantage.

Citedness

The second measure of citation advantage refers to whether articles were cited at all, by individuals other than the authors, during the first 2 years after publication (i.e. over a fixed citation window of 2 years). Since this measure cannot be normalised, it is necessary to use a fixed citation window during which all citations will have been indexed. We chose a 2-year window as the longest window that still offers complete data. We refer to this measure as *citedness*, as opposed to uncitedness (Schwartz 1997), which has been investigated and discussed in other contexts as an indication of research with little to no impact (e.g. Garousi and Fernandes 2017; Cronin, 1984, p. 32), although not without dispute (Mohammed et al. 2020, pp. 1796–1797, Van Noorden 2017). We treat citedness as an indication that an article has been formally incorporated into its broader subject area and has become visible to the research community (Mohammed, Morgan, and Nyantaki 2020, p. 1797). Self-citations (i.e. when a citing article and cited article have at least one author in common, with exactly the same name and surname) were excluded from this measure, as they do not reflect visibility beyond the authors.

To determine whether OA journal articles experience a citation advantage according to this measure, we again consulted the effect size of the relationship between the measure and the variable access status. As analysis is dichotomous-by-dichotomous, the phi-coefficient (φ) was used. A result of $|\varphi| \geq 0.1$ is interpreted as an effect size that is large enough to indicate a relationship, and $|\varphi| \geq 0.3$ is taken to indicate a strong relationship. In each case, a negative value indicates an OA citation advantage.

Most frequently cited

The third measure of citation advantage that was investigated in this study is the percentage of articles among OA journal articles and subscription journal articles that, on the basis of the NCS, belong to the 1%, 5% and 10% most frequently cited articles in their subject area. This measure is often referred to as 'percentage of publications in top 1/5/10%', and in this article it is abbreviated as pptop1/5/10. As this measure is based on the NCS, a 6-year citation window is also applicable to this measure and self-citations are included. Due to

the complications caused by multiple articles potentially having the same number of citations, various methods have been proposed in the literature for calculating these percentiles (Waltman and Schreiber 2013, pp. 373–376). The method chosen for this study was the one applied in the SCImago Institutions Rankings (Waltman and Schreiber 2013, p. 374), according to which all articles that have sufficient number of citations to qualify for the 1% most frequently cited articles are included under the pptop1; and likewise, for 5% and 10%. This method generally results in the inclusion of a percentage of articles higher than X in the pptopX (e.g. more than 1% of articles are included within the 1% most frequently cited articles), but avoids reintroducing the problems associated with a skewed distribution that percentile-based indicators were initially designed to overcome. To determine whether OA articles experience a citation advantage according to this measure, we applied the same statistical procedure as we did for citedness, i.e. φ.

Defining access status

There are many different types of OA, and importantly, these differ in terms of citation behaviour (Archambault et al. 2014b). Conceptual clarity on how articles are classified as OA in the WoS database is also lacking. As a result, the distinction between the different types of OA is not clear-cut. We therefore decided to focus on two types of articles that have the strictest and clearest definitions. The first type, subscription journal articles, we define as those that have no (OA) label in the WoS metadata. They were published in subscription only journals, and no self-archived versions of those articles are available.

The second type, OA journal articles, we define as those published in OA journals listed on the DOAJ. The WoS metadata contains five categories of OA, but these are not mutually exclusive, and therefore articles can be assigned more than one of these categories. One of these categories is Gold DOAJ, which is applied only to those articles published in DOAJ listed journals, and, for the purpose of our study, we classified all articles in this category as OA journal articles. All articles that have one of the other OA labels, but not Gold DOAJ, were excluded on the basis that their definition is unclear, or they refer to self-archived articles. An example of the ambiguity is the Gold Other label: an article with this label may be an OA journal article published in a journal not listed in the DOAJ, or it may be a hybrid OA article. Limiting our definition of OA journal articles to Gold DOAJ led to an exclusion of 22.8% of all the articles indexed for the years investigated. Of all the articles indexed for the years investigated 9.2% are Gold DOAJ articles and 67.9% are subscription journal only articles.

In summary, we define the variable access status as a dichotomous one, comprised of two attributes: Gold OA journal articles published in DOAJ listed journals; and subscription journal articles of which no self-archived versions are available. These two categories are henceforth referred to as 'OA journal article' and 'subscription journal article'.

Data extraction and definition of the population

The data required for this study were extracted from what was, at the time of the study, the most recently available (2018) version of the WoS metadata housed at the Centre for Research on Evaluation, Science and Technology (CREST). The population was defined as all English-language articles and reviews published between 2013 and 2015 in either OA journals or subscription journals, and, in the case of the latter, for which no OA versions were available. This definition and the aim of the study imply selection in terms of

language, document type, publication year, and subject area. The decisions underlying our selection in terms of each of these variables are elaborated upon below.

Publication year

Our selection of the publication years 2013 to 2015 reflects a weighing up of various considerations. First, we needed to include articles in our analysis that were published as recently as possible, in order to account (and compromise) for changes in journals' policy on OA, and how these are managed in WoS. In short, an article published in a journal that is not classified as an OA journal at the time of that article's publication in the journal, is retrospectively classified as an OA journal in WoS if that journal changes its policy to become an OA journal. Thus, an article classified as an OA journal article in the 2018 version of the WoS metadata, may not have been OA when it was originally published and, more importantly, when it received its citations. In order to reduce the likelihood of reproducing retrospective classifications in our data set, we reduced our publication window by excluding articles published before 2013.

We only included articles published up until 2015 to ensure a sufficiently large citation window to calculate the NCS, especially considering our aim to disaggregate by subject area. Thus, to balance the requirements of having sufficiently recent articles (limiting misclassification and exclusion of articles), sufficient number of articles per subject area, and sufficient number of years for articles to receive citations (sufficiently long citation windows), articles published during the three-year period of 2013, 2014, and 2015 were selected and data were extracted from the most recently available (2018) version of the WoS metadata.

Document type

Various document types are indexed in the WoS database (Thomson Reuters 2018, p. 27), and document types published in academic journals differ in the number of citations they tend to receive (Tahamtan et al. 2016, p. 1200). We decided to include both research articles and review articles, as we wanted to focus the study on empirical research articles, however the conceptual distinction between these two document types in WoS is too arbitrary to be operationally useful (Basson 2019, pp. 69–70)—an observation supported by the fact that various previous studies have grouped similar document types together for the purpose of citation analysis (e.g. Van Leeuwen et al. 2001). In addition, including both these similar document types ensured that sufficient number of articles for each subject area were included in the population for analysis.

Subject area

At the time of data extraction WoS used a subject area categorisation scheme that includes 251 subject area labels. Of special note is the subject area label 'Multidisciplinary sciences'. While previous studies that investigated the OA citation advantage reported results on this 'subject area' (Dorta-González et al. 2017; Sotudeh and Horri 2009), we argue that it should be excluded in analyses that specifically investigate subject areas. Our reasoning being that one disaggregates by subject area because citation behaviour differs among subject areas. However, articles with the 'Multidisciplinary sciences' label are those articles published in multidisciplinary journals that include articles from various subject areas. As

such, 'Multidisciplinary sciences' is not a subject area in itself, but rather a description of the type of journal the article is published in. WoS assigns articles published in these multidisciplinary journals with other subject area labels, e.g. based on the article's cited references, when possible (Thomson Reuters 2018, p. 6). Due to these considerations we exclude 'Multidisciplinary sciences' from our analysis of subject areas.

Language

Lastly, included in our study is the research question whether articles published in OA journals are more likely to be written in a language other than English, in other words to determine whether language is a confounding variable when investigating the OA citation advantage of OA journal articles. Previous studies have shown that English-language articles tend to receive more citations than articles published in other languages (Torres-Salinas et al. 2016, pp. 1206–1207; Van Leeuwen et al. 2001, p. 345). In addition, there is some anecdotal evidence that OA journal articles are more likely to be written in a language other than English (Björk 2011, p. e20961).

Previously, an insufficient proportion of articles indexed in WoS were assigned the relevant metadata to test whether there is a relationship between whether an article is published in an OA journal and whether it is written in a language other than English (Basson 2019). In the most recent metadata this has been rectified and now all WoS articles have language data. In this study, we investigated the abovementioned relationship for all the articles published from 2013 to 2015, regardless of subject area. To determine whether there is a relationship between access status and language, we applied the same statistical procedure as we did for citedness to determine the effect size of the relationship. Based on the results of this analysis, which is also reported below, language was determined to be a confounding effect, and only English articles were included for the analysis by subject area.

Results

First, we present the analysis of language as a confounding variable as the results thereof determined the population for the rest of the analysis. Thereafter we describe the prevalence of OA journal articles, in general and by subject area. Specific attention was given to those subject areas with very few OA journal articles as measures of effect size may indicate a relationship for these when none exist due to the small population. We also noted those subject areas with a high percentage of OA journal articles for further discussion. Thereafter we present the analysis of the three measures of citation advantage.

Analysis of language as a potentially confounding variable

For the years 2013 to 2015, 3,902,547 subscription journals and OA journals articles were indexed in WoS at the time when we extracted the data.[1] The vast majority of these

---

[1] A total of 5057,345 reviews and articles are indexed in WoS for the years 2013 to 2015. However, all the analyses in this study, including the analysis of language as a potential confounding variable, pertain to a subset of these, namely subscription journals and OA journals articles, as indicated in the section "Defining access status".
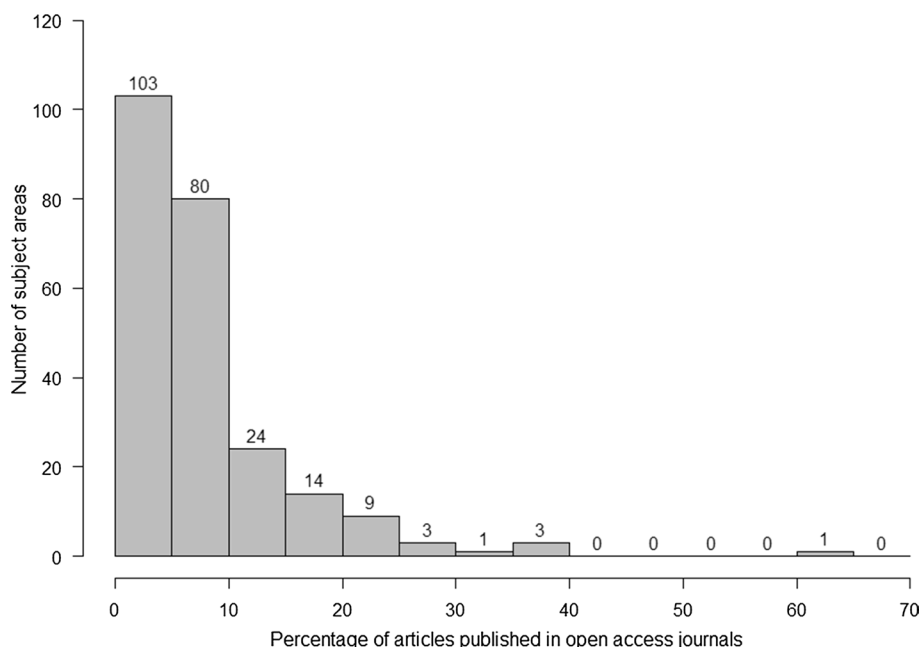
Fig. 1 Number of subject areas by percentage of articles published in open access journals

(94.1%) are written in the English language. Our analysis found that OA journal articles are more likely than subscription journal articles to be written in a language other than English ($\varphi = 0.16$). In order to control for the potentially confounding effect of language, our results presented below were therefore produced by limiting our analyses to English-language articles only.

The prevalence of OA journal articles

Limiting the articles to those written in the English-language resulted in 3,670,682 articles of which 87.3% are published in subscription journals with the remaining 12.7% published in OA journals. The prevalence of OA journal articles in comparison to subscription journal articles for the individual subject areas differ considerably from this. The study investigated 250 subject areas and at the one extreme, we identified 12 subject areas with no OA journal articles published from 2013 to 2015. These subject areas are: 'Education, special', 'Engineering, ocean', 'Ergonomics', 'Ethnic studies', 'Literary reviews', 'Literature, African, Australian, Canadian', 'Literature, American', 'Literature, Slavic', 'Materials science, characterization & testing', 'Physics, fluids & plasmas', 'Psychology, mathematical', and 'Psychology, psychoanalysis'. The majority of the remaining 238 subject areas also tend to have very small percentages of OA journal articles in comparison to subscription journal articles.

The skewness of the distribution is illustrated in Fig. 1 and by the fact that for 183 subject areas (more than three-quarters, or 77% of the 238 with any OA journal articles), less than 10% of articles were published in OA journals. Of the remaining

55 subject areas, the percentage of OA journal articles ranges between 10 to 20% ($10\% \leq x < 20\%$) for the majority (38) of them, while a further 12 fall within the 20–30% range ($20\% \leq x < 30\%$). Only 4 subject areas fall within a range of 30–40% ($20\% \leq x < 40\%$) of OA articles. In order from lowest to highest percentage, these are: 'Integrative & complementary medicine' (32.2%), 'Parasitology' (35,2%), 'Agricultural economics & policy' (36.4%) and 'Medicine, general & internal' (39.8%). Lastly, 'Tropical medicine' (61.4%) is the only subject area with a greater percentage of articles published in OA journals than in subscription journals.

Analysis on the three measures of citation advantage, by subject area

The first measure of citation advantage that we consider is NCS, which we found is related to access status in a relatively small number (76, or 30%) of the 250 subject areas. However, in only a single subject area, 'Art' ($r_{pb} = -0.10$), do OA journal articles experience a citation advantage in comparison to subscription journal articles. In the remaining 75 of these 76 subject areas, the articles in subscription journals show a citation advantage for the measure NCS. Only 'Agricultural economics & policy', has an effect size ($r_{pb} = 0.32$) sufficiently large to indicate a strong association.

Fewer than half (98) of the subject areas show a relationship between our second measure, citedness, and access status, and in the case of only 4 of these subject areas an advantage for OA journal articles was found. For the remaining 94 subject areas, the citation advantage is for subscription journal articles. Amongst these subject areas we found three subject areas for which the effect size is sufficiently large ($\varphi \leq 0.30$) to indicate a strong association.

Lastly, we consider the relationship between access status and each of the three pptop1/5/10 measures. The first observation is that the number of subject areas for which such a relationship was found, is much smaller for this set of measures: 26 for the pptop10, 7 for the pptop5, and only 1 for the pptop1. Secondly, and similar to what we observed for the other measures, the citation advantage tends to be in favour of subscription journal articles, not OA journal articles. In only two of the 26 subject areas for which we found a relationship between access status and pptop10, the citation advantage accrued to OA journal articles (these being 'Art' and 'Architecture'). In only one of the 7 subject areas for which we found a relationship between access status and pptop5, the citation advantage accrued to OA journal articles (this being 'Art'). Deviating slightly from the pattern observed until now, are the results for the last measure: only one subject area experience a citation advantage namely 'Art' for which the advantage is in favour of OA journal articles ($\varphi = -0.10$).

In summary, across all the measures, only 6 of the 250 subject areas in WoS experienced an OA citation advantage for OA journal articles in comparison to subscription journal articles. These subject areas are listed in Table 1, together with the effect sizes obtained for the three measures. Four of the subject areas experience an OA citation advantage on only one measure, namely citedness. These subjects are 'Andrology', 'Engineering, petroleum', 'Physics, nuclear', and 'Tropical medicine'. 'Art' was observed as experiencing an OA citation advantage for the measures NCS, pptop1, pptop5, and pptop10, whereas OA journal articles in 'Architecture' only experience it for the measure pptop10. Table 2 in the appendix shows the results for all the subject areas in WoS.

Table 1 Effect size for each measure of citation advantage for the subject areas that experienced an open access citation advantage

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Art	6221	4.9	−0.12	−0.09	−0.10	−0.12	−0.10
Architecture	5497	8.7	−0.04	−0.07	0.00	−0.09	−0.10
Andrology	1021	25.8	0.00	−0.11	0.04	0.00	0.01
Engineering, petroleum	5788	6.7	0.00	−0.11	0.03	0.02	0.01
Physics, nuclear	16,092	9.5	−0.01	−0.11	0.01	−0.01	−0.02
Tropical medicine	5526	61.4	−0.04	−0.13	0.03	0.05	0.02

Discussion

During the almost two decades since the BOAI declaration, substantial strides have been made in providing OA to academic literature. Numerous studies have investigated the prevalence of OA publications (e.g. Torres-Salinas et al. 2016, p. 19; Dorta-González et al. 2017, p. 880, Archambault et al. 2014a, p. 30). From these we can conclude that in 2009 approximately 5% of all journal articles were published in OA journals and by 2014 this had increased to 10% (Torres-Salinas et al. 2016, p. 19, Dorta-González et al. 2017, p. 880). Our study found that, while less than 10% of all articles and reviews are rendered OA immediately, through publication in DOAJ listed journals, more than 30% of all reviews and articles indexed in the WoS database for the years 2013 to 2015 are rendered OA through various methods. Thus, while OA journal publications still represent only a small proportion of all publications, they constitute a significant proportion of all the OA publications indexed on WoS. However, we encourage caution when using the WoS OA categories to report on the prevalence of OA publications across years. At most one can report what the situation at the time of data extraction was, as in many cases the labels are applied retroactively and as such do not reflect the access status of articles at the time of their publication. Our analysis of prevalence by subject area determined that, for the vast majority of subject areas in the WoS database, at least one OA journal article was indexed for the publication years from 2013 to 2015 in 2018. However, for the majority of the remaining subject areas fewer than 10% of the articles are published in OA journals, with only one subject area in which OA journal publishing was more common than subscription journal publishing.

Regarding the potential confounding variable language, the anecdotal assertion that OA journal articles are more likely to be written in languages other than English was supported by our study and, in turn, supports our decision to limit our analyses to English articles and reviews. This finding has methodological implications for future research, as it leads us to recommend controlling for language, where possible, especially since WoS has improved the availability of language data for indexed articles (Basson 2019, p. 71). The next step would be to investigate how subject areas differ in terms of the extent to which OA journal articles are published in a language other than English—an issue that unfortunately fell beyond the possible scope of this article.

As to whether OA journal articles experience a citation advantage, it has been reported elsewhere (e.g. Archambault et al. 2014b, p. 20, Van Leeuwen, Tatum and Wouters 2018, p. 12, Basson 2019, p. 108) that OA journal articles in general do not experience an OA citation advantage, and we also find this to be the case for the vast majority of subject areas. However, in a small number of subject areas OA journal articles do experience a citation advantage, and they do so mostly on the measure of citedness. This may reflect the "*new visibility*" and *"readership"* that, according to the BOAI declaration, are provided by rendering an article OA. However, the fact that so few subject areas experience this type of OA citation advantage for OA journal articles in comparison to subscription journal articles points to the need to further investigation why this advantage exists only for those few subject areas.

Previously Dorta-González et al. (2017) also investigated whether OA journal articles in the various subject areas defined by WoS, experience a citation advantage, using the WoS database in a similar way than we did. Dorta-González et al. used "*proportion of the average citation of OA articles in relation to non-OA articles*" (Dorta-González et al. 2017, p. 880) as their measure of the OA citation advantage. This measure is somewhat comparable to the NCS measure used in our study. Their study differs from ours in that they did not control for language, investigated a different period (2009–2014), considered any positive value of the proportion of the average citation of OA journal articles in relation to non-OA journal articles as indicating a citation advantage (without considering effect size), and did not compare OA journal articles with subscription journal articles, but with all non-OA journal articles, using the previous WoS OA categorisation system. However, some comparison between their results and ours is useful. While we found only a single subject area ('Art') in which OA journal articles experienced a citation advantage on this measure, they found this for a larger number of subject areas (36 out of 249 for which 'Art' experienced the strongest advantage), but in the majority of subject areas (173) they find an OA citation disadvantage on this measure.

Clearly, factors other than whether an article is published in an OA journal or a subscription journal need to be taken into account when trying to explain variation in citation. While we already discussed language, another factor is suggested in the BOAI declaration when it states that providing OA to academic research would lead to the proclaimed benefits. This suggests a potential relationship between prevalence and OA citation advantage. 'Tropical medicine' at 61%, is the only subject area for which the majority of articles are published in OA journals and OA journal articles experienced a citation advantage for one of the measures (citedness). However, in 'Art' the prevalence of OA journal articles is particularly low (4.9%) and OA journal articles experience a citation advantage for multiple measures. In the case of 'Medicine, general & internal' a significant proportion (39.8%) of articles are OA journal articles, but subscription journal articles experience a citation advantage for four of the measures. On the other hand, several of the subject areas that do experience an OA citation advantage have a low prevalence of OA journal articles (< 10%). It would seem that attaining the benefits suggested in the BOAI declaration is not simply a matter of increasing prevalence of OA journal publications, at least in the case of citation advantage for OA journal articles, when compared to subscription journal articles. However, the potential relationship between OA journal article prevalence and the measures of citation advantage was not tested in this study.

Our results show that other factors have a stronger effect on the number of citations that an article receives, than whether the journal in which it is published applies an OA publishing model. We reiterate that our definition of what constitutes an OA journal article is limited to those articles published in a DOAJ-listed journal, and not to all types of OA

e.g. self-archived articles or hybrid OA articles. These DOAJ journals might systematically differ from other journals that allow OA to their publications, which leads us to strongly recommend taking into account additional journal-related factors in future investigations of the OA citation advantage. We further recommend that future research on the OA citation advantage investigate more factors related to the journal characteristics beyond the access policies of the journals, such as language, prevalence of OA publication, prestige of the journal, and the publication and review policies of journals e.g. OA megajournals. Other article and author related factors, such as collaboration types, could also be considered as potentially relevant variables. In addition, we recommend that future studies include measures of effect size and explicit descriptions of the access types, subject areas, and document types included in the study population to assist with the comparability of future studies. Finally, our study highlights that the WoS OA label functions differently from most other data in the WoS metadata, due to its retroactive application, and this characteristic should be considered in future research aiming to use the WoS OA label.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Appendix

See Table 2.

Table 2 Effect size for each measure of citation advantage for each of the subject areas of WoS

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Acoustics	11,347	5.2	0.12	0.14	0.02	0.04	0.06
Agricultural economics & policy	3037	36.4	0.32	0.36	0.08	0.16	0.21
Agricultural engineering	10,884	5.8	0.20	0.34	0.03	0.06	0.09
Agriculture, dairy & animal science	15,933	12.9	0.14	0.14	0.03	0.06	0.09
Agriculture, multidisciplinary	16,611	11.0	0.16	0.17	0.04	0.08	0.12
Agronomy	23,285	11.9	0.16	0.19	0.03	0.08	0.11
Allergy	5237	8.1	0.11	0.07	0.03	0.06	0.08
Anatomy & morphology	3943	3.9	− 0.04	− 0.08	0.02	0.00	− 0.02
Andrology	1021	25.8	0.00	− 0.11	0.04	0.00	0.01
Anesthesiology	10,999	16.6	0.20	0.28	0.04	0.09	0.14
Anthropology	10,458	2.0	0.03	0.05	0.01	0.02	0.02
Archaeology	6975	5.8	0.12	0.15	0.03	0.06	0.08
Architecture	5497	8.7	− 0.04	− 0.07	0.00	− 0.09	− 0.10
Area studies	8708	7.6	0.10	0.11	0.03	0.06	0.07
Art	6221	4.9	− 0.12	− 0.09	− 0.10	− 0.12	− 0.10
Asian studies	3785	6.4	0.05	0.04	0.02	0.02	0.04
Astronomy & astrophysics	25,557	6.4	0.00	0.02	0.00	0.00	− 0.01
Audiology & Speech-Language Pathology	4831	3.5	0.07	0.03	0.01	0.03	0.05
Automation & control systems	25,114	2.0	0.06	0.12	0.01	0.03	0.05
Behavioral Sciences	13,928	4.9	0.09	0.04	0.02	0.05	0.07
Biochemical research methods	31,287	9.0	0.04	0.06	0.02	0.04	0.04
Biochemistry & molecular biology	91,730	10.0	0.07	0.00	0.02	0.05	0.06
Biodiversity conservation	12,337	8.4	0.12	0.10	0.03	0.06	0.09
Biology	18,653	20.3	0.04	0.00	0.03	0.02	0.02
Biophysics	25,755	0.6	0.03	0.04	0.01	0.02	0.02

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Biotechnology & applied microbiology	56,717	16.8	0.05	0.08	0.00	0.04	0.07
Business	23,570	3.7	0.09	0.13	0.02	0.04	0.06
Business, finance	17,381	2.2	0.07	0.08	0.01	0.03	0.05
Cardiac & cardiovascular systems	37,738	11.0	0.11	0.16	0.02	0.05	0.07
Cell & Tissue Engineering	3770	19.6	0.07	0.05	0.03	0.04	0.06
Cell biology	34,592	15.5	0.07	0.04	0.03	0.06	0.07
Chemistry, analytical	60,908	4.8	0.05	0.02	0.02	0.03	0.04
Chemistry, applied	39,551	1.4	0.08	0.11	0.01	0.03	0.04
Chemistry, inorganic & nuclear	37,094	0.4	0.03	0.02	0.01	0.02	0.02
Chemistry, medicinal	35,287	4.5	0.04	0.00	0.02	0.03	0.04
Chemistry, multidisciplinary	156,532	5.9	0.08	0.08	0.03	0.06	0.08
Chemistry, organic	52,133	1.4	0.04	0.03	0.01	0.02	0.03
Chemistry, physical	162,339	0.5	0.03	0.06	0.01	0.02	0.02
Classics	1791	7.0	−0.07	−0.02	−0.08	−0.06	−0.04
Clinical neurology	52,719	6.8	0.10	0.13	0.02	0.05	0.07
Communication	12,032	5.7	0.09	0.14	0.02	0.05	0.07
Computer science, artificial intelligence	35,500	2.6	0.06	0.10	0.02	0.04	0.05
Computer science, cybernetics	4420	2.9	0.08	0.15	0.02	0.04	0.06
Computer science, hardware & architecture	14,484	0.4	0.03	0.07	0.01	0.02	0.02
Computer science, information systems	39,099	8.3	0.10	0.15	0.03	0.07	0.09
Computer science, interdisciplinary applications	35,993	2.2	0.05	0.10	0.01	0.03	0.04
Computer science, software engineering	22,989	2.6	0.05	0.09	0.02	0.03	0.04
Computer science, theory & methods	22,732	4.0	0.07	0.20	0.02	0.05	0.07
Construction & building technology	19,968	1.9	0.07	0.08	0.01	0.03	0.04
Criminology & penology	7320	2.0	0.06	0.05	0.02	0.03	0.04

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Critical care medicine	10,302	7.0	0.09	0.20	0.03	0.06	0.08
Crystallography	20,596	20.5	0.17	0.42	0.05	0.11	0.16
Cultural Studies	4773	3.7	0.04	0.10	0.02	0.03	0.06
Dance	665	0.2	0.01	0.01	0.01	0.01	−0.02
Demography	2768	14.6	0.08	0.06	0.04	0.05	0.08
Dentistry, oral surgery & medicine	24,218	8.8	0.12	0.15	0.03	0.06	0.08
Dermatology	17,251	9.7	0.11	0.11	0.03	0.05	0.07
Developmental biology	5026	3.2	0.01	−0.04	0.01	0.01	0.02
Ecology	38,741	6.9	0.07	0.01	0.02	0.05	0.06
Economics	56,616	4.4	0.09	0.13	0.02	0.04	0.06
Education & educational research	44,638	9.5	0.08	0.08	0.03	0.06	0.07
Education, Scientific Disciplines	10,154	7.3	−0.01	0.03	0.01	0.01	−0.01
Education, special	4859	0.0	NA	NA	NA	NA	NA
Electrochemistry	40,002	6.5	0.09	0.07	0.03	0.05	0.07
Emergency medicine	9327	11.2	0.09	0.07	0.03	0.06	0.07
Endocrinology & metabolism	31,293	9.6	0.10	0.09	0.02	0.05	0.07
Energy & fuels	79,527	3.0	0.08	0.09	0.02	0.04	0.06
Engineering, aerospace	9029	7.5	0.01	−0.03	0.02	0.02	0.01
Engineering, biomedical	27,910	2.5	0.05	0.06	0.02	0.03	0.04
Engineering, chemical	83,875	1.5	0.06	0.08	0.01	0.03	0.04
Engineering, civil	42,902	2.2	0.08	0.11	0.02	0.03	0.04
Engineering, electrical & electronic	141,060	3.2	0.07	0.11	0.02	0.04	0.05
Engineering, environmental	34,094	0.1	0.02	0.02	0.00	0.01	0.01
Engineering, geological	9753	1.5	0.02	0.01	0.01	0.01	0.03
Engineering, industrial	14,134	3.8	0.12	0.20	0.02	0.05	0.07

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Engineering, manufacturing	17,167	0.5	0.04	0.06	0.01	0.02	0.02
Engineering, marine	2758	20.5	0.23	0.27	0.05	0.09	0.15
Engineering, mechanical	52,184	5.5	0.12	0.17	0.02	0.05	0.07
Engineering, multidisciplinary	38,515	20.8	0.17	0.23	0.05	0.10	0.15
Engineering, ocean	3871	0.0	NA	NA	NA	NA	NA
Engineering, petroleum	5788	6.7	0.00	−0.11	0.03	0.02	0.01
Entomology	15,475	7.8	0.10	0.11	0.03	0.06	0.08
Environmental Sciences	106,248	3.8	0.05	0.04	0.02	0.04	0.05
Environmental studies	20,555	8.5	0.12	0.04	0.03	0.07	0.10
Ergonomics	4230	0.0	NA	NA	NA	NA	NA
Ethics	6202	5.0	0.04	0.05	0.01	0.01	0.03
Ethnic studies	2838	0.0	NA	NA	NA	NA	NA
Evolutionary biology	9556	15.3	0.12	−0.02	0.04	0.08	0.11
Family studies	6691	1.3	0.07	0.10	0.01	0.02	0.03
Film, radio, television	3196	3.9	0.07	0.10	0.02	0.05	0.07
Fisheries	13,253	5.3	0.08	0.04	0.02	0.05	0.06
Folklore	424	9.9	0.01	−0.01	−0.01	0.03	0.05
Food science & technology	58,448	5.6	0.14	0.21	0.02	0.05	0.08
Forestry	11,100	9.0	0.13	0.14	0.03	0.05	0.08
Gastroenterology & hepatology	24,354	9.1	0.14	0.16	0.03	0.06	0.09
Genetics & heredity	32,246	20.8	−0.02	−0.02	0.00	0.00	0.01
Geochemistry & geophysics	22,396	3.2	0.08	0.08	0.02	0.03	0.04
Geography	12,539	7.4	0.14	0.23	0.03	0.06	0.09
Geography, physical	14,634	4.7	0.07	0.05	0.02	0.04	0.05
Geology	5842	8.7	0.16	0.14	0.03	0.08	0.11

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Geosciences, multidisciplinary	53,750	8.4	0.05	0.01	0.02	0.04	0.05
Geriatrics & Gerontology	9718	8.9	0.05	−0.02	0.02	0.05	0.05
Gerontology	5435	2.8	0.02	−0.02	0.01	0.03	0.03
Health care sciences & services	19,057	18.9	0.12	0.13	0.03	0.07	0.08
Health policy & services	12,039	10.0	0.04	0.00	0.02	0.05	0.06
Hematology	17,885	7.4	−0.04	0.10	0.00	−0.02	−0.05
History	16,528	3.6	0.06	0.06	0.02	0.04	0.04
History & Philosophy Of Science	6366	5.4	0.10	0.13	0.02	0.05	0.07
History of social sciences	2916	0.6	0.04	0.07	0.01	0.02	0.03
Horticulture	9146	3.4	0.08	0.09	0.02	0.04	0.05
Hospitality, Leisure, Sport & Tourism	9987	9.4	0.18	0.23	0.03	0.08	0.11
Humanities, multidisciplinary	12,903	12.1	0.05	0.06	0.02	0.04	0.05
Imaging science & photographic technology	10,432	1.3	0.05	0.08	0.01	0.03	0.03
Immunology	34,668	11.6	0.10	0.07	0.03	0.06	0.08
Industrial relations & labor	3375	7.1	0.03	−0.04	0.03	0.05	0.06
Infectious diseases	21,542	22.9	0.10	0.08	0.04	0.08	0.09
Information science & library science	12,399	7.1	0.10	0.14	0.03	0.06	0.08
Instruments & instrumentation	39,916	8.6	0.07	−0.04	0.03	0.06	0.07
Integrative & complementary medicine	9169	32.2	0.17	0.18	0.05	0.08	0.10
International relations	11,207	2.3	0.08	0.12	0.02	0.04	0.05
Language & Linguistics	12,172	9.7	0.12	0.13	0.04	0.08	0.09
Law	21,612	2.5	0.04	0.05	0.01	0.03	0.04
Limnology	3667	3.9	0.04	−0.01	0.01	0.03	0.04
Linguistics	13,593	4.4	0.07	0.09	0.02	0.03	0.04
Literary reviews	2680	0.0	NA	NA	NA	NA	NA

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Literary theory & criticism	1698	8.4	0.04	0.06	0.03	0.03	0.02
Literature	10,174	5.4	0.05	0.08	0.03	0.05	0.07
Literature, African, Australian, Canadian	645	0.0	NA	NA	NA	NA	NA
Literature, American	1490	0.0	NA	NA	NA	NA	NA
Literature, British Isles	1176	5.4	0.12	0.11	0.03	0.06	0.09
Literature, German, Dutch, Scandinavian	561	0.2	0.02	0.02	0.01	0.02	0.02
Literature, Romance	1352	1.0	0.03	0.00	0.02	0.04	0.02
Literature, Slavic	228	0.0	NA	NA	NA	NA	NA
Logic	2628	9.6	0.07	0.11	−0.02	−0.01	0.04
Management	33,075	4.7	0.12	0.18	0.02	0.05	0.07
Marine & freshwater biology	26,495	3.2	0.06	0.04	0.01	0.03	0.04
Materials science, biomaterials	18,538	0.8	0.05	0.08	0.01	0.02	0.03
Materials science, ceramics	14,865	3.3	0.08	0.08	0.02	0.04	0.05
Materials science, characterization & testing	7424	0.0	NA	NA	NA	NA	NA
Materials science, coatings & films	18,997	0.4	0.01	0.02	0.00	0.00	0.00
Materials science, composites	11,113	0.4	0.04	0.06	0.01	0.02	0.02
Materials science, multidisciplinary	220,481	4.4	0.08	0.11	0.02	0.05	0.06
Materials science, paper & wood	5478	23.0	0.07	−0.02	0.05	0.09	0.09
Materials science, textiles	6792	4.3	0.08	0.10	0.02	0.05	0.06
Mathematical & Computational Biology	11,708	28.1	0.09	0.01	0.03	0.08	0.09
Mathematics	73,858	14.5	0.11	0.10	0.03	0.06	0.09
Mathematics, applied	69,067	13.9	0.14	0.19	0.03	0.07	0.10
Mathematics, interdisciplinary applications	25,834	21.8	0.19	0.28	0.05	0.11	0.16
Mechanics	52,006	2.4	0.08	0.10	0.02	0.03	0.05
Medical ethics	2032	13.5	0.03	0.08	0.01	0.00	0.03

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Medical informatics	5018	10.4	−0.01	−0.05	0.01	0.03	0.01
Medical laboratory technology	7072	3.2	0.10	0.15	0.01	0.03	0.05
Medicine, general & internal	50,406	39.8	0.10	0.03	0.07	0.14	0.17
Medicine, legal	5430	0.9	0.00	0.07	−0.01	−0.02	−0.01
Medicine, research & experimental	40,392	26.7	0.14	0.11	0.04	0.08	0.11
Medieval & renaissance studies	2137	8.0	0.08	0.07	0.02	0.05	0.08
Metallurgy & metallurgical engineering	41,967	5.2	0.09	0.11	0.02	0.05	0.07
Meteorology & Atmospheric Sciences	27,697	14.7	0.07	−0.01	0.03	0.06	0.07
Microbiology	23,770	15.1	0.05	0.00	0.03	0.05	0.05
Microscopy	2766	0.3	0.00	0.00	0.00	0.01	−0.01
Mineralogy	7440	1.1	−0.01	0.02	0.01	−0.02	−0.01
Mining & mineral processing	6624	5.8	0.10	0.12	0.03	0.04	0.06
Music	4087	3.6	0.05	0.06	0.01	0.01	0.04
Mycology	4109	7.8	0.11	0.15	0.02	0.05	0.07
Nanoscience & nanotechnology	83,651	7.1	0.10	0.12	0.03	0.06	0.09
Neuroimaging	4261	4.7	0.05	−0.01	0.02	0.04	0.06
Neurosciences	65,214	11.1	0.09	0.06	0.03	0.06	0.07
Nuclear science & technology	26,661	2.2	0.03	0.06	0.00	0.02	0.03
Nursing	22,796	9.0	0.12	0.17	0.03	0.07	0.09
Nutrition & dietetics	20,376	8.5	0.12	0.05	0.03	0.06	0.09
Obstetrics & gynecology	26,577	7.2	0.10	0.11	0.02	0.05	0.07
Oceanography	14,775	5.6	0.09	0.04	0.01	0.04	0.06
Oncology	66,538	16.0	0.12	0.08	0.03	0.07	0.10
Operations research & management science	23,862	0.4	0.04	0.09	0.01	0.02	0.02
Ophthalmology	18,297	18.6	0.16	0.17	0.04	0.10	0.14

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Optics	72,447	13.2	−0.02	−0.08	0.02	−0.01	−0.03
Ornithology	2973	1.8	0.05	0.00	0.01	0.03	0.04
Orthopedics	28,418	12.8	0.15	0.14	0.03	0.07	0.10
Otorhinolaryngology	14,279	6.1	0.11	0.15	0.02	0.05	0.07
Paleontology	6963	4.4	0.05	0.04	0.01	0.02	0.02
Parasitology	10,912	35.2	0.05	0.00	0.05	0.07	0.07
Pathology	19,843	9.1	0.11	0.11	0.02	0.05	0.08
Pediatrics	38,077	8.1	0.12	0.19	0.02	0.04	0.07
Peripheral vascular disease	17,314	3.9	0.07	0.07	0.01	0.03	0.04
Pharmacology & pharmacy	84,266	7.0	0.09	0.10	0.02	0.05	0.07
Philosophy	15,786	7.9	0.09	0.10	0.03	0.06	0.08
Physics, applied	152,780	3.1	0.04	0.04	0.02	0.04	0.05
Physics, atomic, molecular & chemical	44,117	1.2	0.05	0.07	0.01	0.02	0.03
Physics, condensed matter	75,854	2.1	0.05	0.13	0.01	0.03	0.05
Physics, fluids & plasmas	22,778	0.0	NA	NA	NA	NA	NA
Physics, mathematical	26,953	1.7	0.05	0.06	0.01	0.03	0.04
Physics, multidisciplinary	54,158	5.9	0.00	−0.07	0.01	0.01	0.00
Physics, nuclear	16,092	9.5	−0.01	−0.11	0.01	−0.01	−0.02
Physics, particles & fields	26,672	19.6	0.01	−0.05	0.02	0.03	0.03
Physiology	19,378	12.0	0.08	0.03	0.02	0.06	0.06
Planning & development	9592	0.4	0.04	0.09	0.01	0.01	0.02
Plant sciences	47,839	7.9	0.07	0.06	0.02	0.04	0.05
Poetry	470	0.4	0.03	0.02	0.01	0.02	0.03
Political science	20,852	2.4	0.07	0.10	0.02	0.04	0.05
Polymer science	50,341	1.4	0.04	0.04	0.01	0.02	0.03

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Primary Health Care	2825	22.0	0.03	0.05	0.01	0.03	0.03
Psychiatry	37,789	6.4	0.10	0.06	0.02	0.05	0.06
Psychology	13,233	8.4	0.09	-0.01	0.03	0.06	0.07
Psychology, applied	10,442	1.6	0.06	0.09	0.01	0.03	0.04
Psychology, biological	3598	0.3	0.03	0.03	0.00	0.01	0.02
Psychology, clinical	17,240	1.8	0.05	0.05	0.01	0.03	0.04
Psychology, developmental	10,070	0.5	-0.01	-0.03	0.01	0.01	-0.01
Psychology, educational	5709	1.8	0.08	0.07	0.01	0.03	0.05
Psychology, experimental	14,861	2.1	0.06	0.07	0.02	0.03	0.05
Psychology, mathematical	1194	0.0	NA	NA	NA	NA	NA
Psychology, Multidisciplinary	19,685	16.2	0.11	0.04	0.04	0.09	0.11
Psychology, psychoanalysis	1537	0.0	NA	NA	NA	NA	NA
Psychology, social	9921	0.4	0.05	0.09	0.01	0.02	0.02
Public administration	5332	4.3	0.12	0.16	0.02	0.04	0.07
Public, Environmental & Occupational Health	54,140	18.0	0.10	0.09	0.04	0.08	0.10
Radiology, nuclear medicine & medical imaging	41,975	9.8	0.09	0.09	0.02	0.05	0.06
Rehabilitation	18,933	6.4	0.10	0.14	0.02	0.03	0.05
Religion	13,466	5.6	0.03	0.04	0.01	0.02	0.02
Remote sensing	11,501	12.8	0.08	0.04	0.04	0.07	0.09
Reproductive biology	8205	4.3	0.08	0.04	0.02	0.04	0.06
Respiratory system	15,636	13.0	0.10	0.13	0.03	0.07	0.09
Rheumatology	8679	16.1	0.12	0.16	0.03	0.07	0.10
Robotics	4985	14.0	0.18	0.29	0.04	0.09	0.13
Social issues	4756	4.9	0.10	0.12	0.02	0.04	0.06
Social sciences, biomedical	6768	2.2	0.01	0.03	0.01	0.01	0.02

Table 2 (continued)

Subject area	Articles (total)	% OA	NCS (r_{pb})	Citedness (φ)	Pptop1 (φ)	Pptop5 (φ)	Pptop10 (φ)
Social sciences, interdisciplinary	18,676	9.8	0.12	0.16	0.03	0.07	0.09
Social sciences, mathematical methods	6242	2.7	0.06	0.12	0.02	0.03	0.04
Social work	8097	0.8	0.06	0.08	0.01	0.02	0.03
Sociology	15,234	3.0	0.08	0.12	0.01	0.03	0.05
Soil Science	11,527	4.8	0.12	0.15	0.02	0.05	0.07
Spectroscopy	20,359	1.8	0.05	0.11	0.01	0.01	0.02
Sport Sciences	23,798	6.7	0.12	0.12	0.03	0.06	0.07
Statistics & probability	21,392	1.8	0.03	0.06	0.00	0.01	0.02
Substance Abuse	5778	5.3	0.07	0.04	0.02	0.05	0.06
Surgery	80,936	5.7	0.11	0.16	0.02	0.05	0.07
Telecommunications	38,438	8.2	0.09	0.18	0.03	0.07	0.09
Theater	1992	0.2	0.01	0.02	0.01	0.01	0.02
Thermodynamics	34,866	5.8	0.14	0.21	0.02	0.05	0.08
Toxicology	24,718	5.0	0.04	0.07	0.02	0.02	0.03
Transplantation	8428	0.6	0.03	0.09	0.01	0.01	0.02
Transportation	9713	2.9	0.08	0.07	0.02	0.04	0.05
Transportation science & technology	12,755	3.1	0.08	0.11	0.02	0.04	0.06
Tropical medicine	5526	61.4	−0.04	−0.13	0.03	0.05	0.02
Urban studies	6903	2.9	0.10	0.22	0.02	0.04	0.06
Urology & nephrology	21,493	14.1	0.12	0.16	0.04	0.08	0.10
Veterinary sciences	31,629	13.6	0.11	0.13	0.03	0.04	0.07
Virology	8980	18.9	−0.05	−0.02	0.03	0.01	−0.01
Water resources	36,095	5.8	0.03	0.00	0.02	0.03	0.03
Women's Studies	4494	0.7	0.05	0.08	0.01	0.02	0.03
Zoology	28,813	9.3	0.06	0.07	0.02	0.03	0.04

References

Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics*. https://doi.org/10.1023/A:10219 19228368.

Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*. https://doi.org/10.5860/crl.65.5.372.

Archambault, E., Amyot, D., Campbell, D., Caruso, J., Deschamps, P., Nicol, A., et al. (2014). Evolution of open access policies and availability, 1996–2013. RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access. Report. Science-Metrix. Retrieved from August 20, 2017 from http://science-metrix.com/files/science-metrix/publications/d_4.5_sm_ec_dg-rtd_oa_policies_proportion_oa_1996-2013_v05p_0.pdf.

Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., et al. (2014). Proportion of open access papers published in peer-reviewed journals at the European and world level—1996–2013. RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access. Report. Science-Metrix. Retrieved January 16, 2020 from https://science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf.

Basson, I. (2019). An investigation of open access citation advantage through multiple measures and across subject areas for articles published from 2005 to 2014. PhD thesis. Stellenbosch: Stellenbosch University. Retrieved January 15, 2020 from http://scholar.sun.ac.za/handle/10019.1/105966.

Björk, B.-C. (2011). A study of innovative features in scholarly open access journals. *Journal of Medical Internet Research*. https://doi.org/10.2196/jmir.1802.

Burrell, Q. L. (2002). Will this paper ever be cited? *Journal of the American Society for Information Science and Technology*. https://doi.org/10.1002/asi.10031.

Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genova, Y., Guédon, J.-C., et al.. (2002). *Read the Budapest open access initiative*. Declaration. Open Society Foundations. Retrieved January 15, 2020 from http://openscience.ens.fr/ABOUT_OPEN_ACCESS/DECLARATIONS/2002_02_14_Budapest_Open_Access_Initiative.pdf.

Clarivate Analytics. (2018). *New features and announcements*. Resource page. Clarivate Analytics. Retrieved July 21, 2020 from https://images.webofknowledge.com/images/help/WOS/hp_whatsnew_wos.html.

Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*. https://doi.org/10.1016/j.joi.2007.04.001.

Cronin, B. (1984). *The citation process*. London: Taylor Graham. https://doi.org/10.1016/0306-4573(85)90080-9.

Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: A study of conflicting paradigms. *Journal of Academic Librarianship*. https://doi.org/10.1016/j.acalib.2011.07.002.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*. https://doi.org/10.1007/s11192-007-1661-8.

Dorta-González, P., González-Betancor, S. M., & Dorta-González, M. I. (2017). Reconsidering the gold open access citation advantage postulate in a multidisciplinary context: An analysis of the subject categories in the Web of Science database 2009–2014. *Scientometrics*. https://doi.org/10.1007/s11192-017-2422-y.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.0040157.

Fukuzawa, N. (2017). Characteristics of papers published in journals: An analysis of open access journals, country of publication, and languages used. *Scientometrics*. https://doi.org/10.1007/s11192-017-2414-y.

Gargouri, Y., Hajjem, C., Lariviére, V., Gingras, Y., Carr, L., Brody, T., et al. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0013636.

Garousi, V., & Fernandes, J. M. (2017). Quantity versus impact of software engineering papers: a quantitative study. *Scientometrics*. https://doi.org/10.1007/s11192-017-2419-6.

Glänzel, W., & Thijs, B. (2004). The influence of author self-citations on bibliometric macro indicators. *Scientometrics*. https://doi.org/10.1023/B:SCIE.0000018535.99885.e9.

Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data Engineering Bulletin, 28*(4), 39–46.

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2008). The access/impact problem and the green and gold roads to open access: An update. *Serials Review*. https://doi.org/10.1016/j.serrev.2007.12.005.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., et al. (2005). The effect of use and access on citations. *Information Processing and Management*. https://doi.org/10.1016/j.ipm.2005.03.010.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*. https://doi.org/10.1038/35079151.

Miguel, S., de Oliveira, E. F. T., & Grácio, M. C. C. (2016). Scientific production on open access: A worldwide bibliometric analysis in the academic and scientific context. *Publications*. https://doi.org/10.3390/publications4010001.

Mohammed, S., Morgan, A., & Nyantaki, E. (2020). On the influence of uncited publications on a researcher's h-index. *Scientometrics*. https://doi.org/10.1007/s11192-020-03356-1.

Perianes-Rodríguez, A., & Olmeda-Gómez, C. (2019). Effects of journal choice on the visibility of scientific publications: A comparison between subscription-based and full Open Access models. *Scientometrics*. https://doi.org/10.1007/s11192-019-03265-y.

Porter, A. L. (1977). Citation analysis: Queries and caveats. *Social Studies of Science, 7*(2), 257–267.

Schwartz, C. A. (1997). The rise and fall of uncitedness. *College and Research Libraries*. https://doi.org/10.5860/crl.58.1.19.

Sotudeh, H., & Horri, A. (2007). The citation performance of open access journals: A disciplinary investigation of citation distribution models. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.20676.

Sotudeh, H., & Horri, A. (2009). Countries positioning in open access journals system: An investigation of citation distribution patterns. *Scientometrics*. https://doi.org/10.1007/s11192-009-1870-4.

Tahamtan, I., Safipour, A. A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*. https://doi.org/10.1007/s11192-016-1889-2.

Thomson Reuters. (2018). *InCites: Indicators handbook*. Resource document. Clarivate Analytics. Retrieved January 15, 2020 from http://help.prod-incites.com/inCites2Live/8980-TRS/version/default/part/AttachmentData/data/InCites-Indicators-Handbook%20-%20June%202018.pdf.

Tijssen, R. J. W., Visser, M. S., & Van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics, 54*(3), 381–397.

Torres-Salinas, D., Robinson-García, N., & Aguillo, I. F. (2016). Bibliometric and benchmark analysis of gold open access in Spain: Big output and little impact. *El profesional de la información*. https://doi.org/10.3145/epi.2016.ene.03.

Van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & Van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics, 51*(1), 335–346.

Van Leeuwen, T. N., Tatum, C., & Wouters, P. F. (2018). Exploring possibilities to use bibliometric data to monitor gold open access publishing at the national level. *Journal of the American Society for Information Science and Technology*. https://doi.org/10.1002/asi.24029.

Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & Van Raan, A. F. J. (2003). The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics, 57*(2), 257–280.

Van Noorden, R. (2017). The science that's never been cited. *Nature*. https://doi.org/10.1038/d41586-017-08404-0.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*. https://doi.org/10.1016/j.joi.2016.02.007.

Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. *Journal of the American Society for Information Science and Technology*. https://doi.org/10.1002/asi.

Waltman, L., Van Eck, N. J., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics*. https://doi.org/10.1007/s11192-011-0354-5.