Check for
updates

The open access citation premium may depend on the openness and inclusiveness of the indexing database, but the relationship is controversial because it is ambiguous where the open access boundary lies

Sergio Copiello[1] (ORCID)

Abstract

Do open access (OA) documents benefit from a citation premium in comparison to traditional subscription-based articles? The question has been debated during the last two decades, as OA is gaining momentum and becoming an ever more established option. Without coming to a shared position, the literature on the topic has essentially split into two clusters: on the one hand, the studies that endorse the occurrence of an OA citation advantage; on the other hand, the works suggesting that it has a negligible extent, or is due to other confounding factors. The primary aim of this study is not to bring new evidence in favor or against the citation premium supposedly characterizing OA articles. Instead, this work is meant to test a specific hypothesis connected with the OA citation advantage, namely, that OA papers may benefit from higher citations in open indexing databases (e.g., Google Scholar) rather than in selective indexing engines (i.e., Scopus and Web of Science). The empirical findings, although conflicting, show that the hypothesis above is not misplaced since a few confirmatory results are achieved. However, the hypothesized relationship remains controversial, also because of an uncertain boundary between OA and paywall articles.

Keywords Publishing industry · Open access · Citation premium · Citation advantage · Indexing databases · Openness and inclusiveness

Foreword

This paper situates itself in the ongoing debate about whether or not open access (OA) articles are more cited than documents published in traditional subscription-based journals. Nonetheless, the main points of this study are not directly about the so-called OA citation advantage but about two related aspects. The first lies in the fact that the disagreement on the actual occurrence of that advantage may be due to the different ways in which OA

✉ Sergio Copiello
sergio.copiello@iuav.it

[1] Department of Architecture, IUAV University of Venice, Dorsoduro 2206, 30123 Venice, Italy

is defined or operationalized. The second is that OA articles may be reaching a distinct audience than subscription-based publications, with a citation premium accruing from non-scholarly or less selective sources.

In just a couple of decades since the term was first coined and circulated (Bernius et al. 2009), and under the framework of an evolving scholarly communication landscape (Odlyzko 2002), OA is gaining momentum and is also becoming an ever more established publishing model (Chi Chang 2006; Laakso et al. 2011; Laakso and Björk 2012, 2016; Wang et al. 2018). The literature delving into the potential impact advantage of OA was already copious and significant during the 2000s, then other studies added further insights into the matter. In the next section, a selection of noteworthy studies is first summarized, then attention is paid to the literature that more closely discusses the issues addressed here. The examined works can be roughly clustered into two groups: those supporting the occurrence of a citation advantage, and those doubting that it takes place or indicating that, if any advantage exists, OA is not the primary cause. A long list of studies on the matter of AO and citation practices is recalled in a recent paper (Haug 2019, see endnote 5) and can be found in the SPARC (Scholarly Publishing and Academic Resources Coalition) Europe webpage (https://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/oaca-list/, accessed 02.04.2019).

The remainder of this article is meant to test a specific hypothesis connected with the OA citation advantage, namely, that OA papers may accrue more citations in open indexing databases than in selective indexing engines. In other words, the additional cites that give birth to the OA citation premium may be unevenly distributed between literature sources (peer-reviewed documents, which are featured in selective indexing databases) and other sources (working papers, pre-print versions, lecture notes, patents, and so forth, which are extensively retrieved by non-selective search engines).

After presenting the method, the selected case studies, and the data gathering process, the results are discussed with two separate focuses on Gold OA articles and Green OA (self-archived) documents, respectively. The empirical findings show that the hypothesis above is not misplaced since some confirmatory results are achieved. However, the hypothesized relationship remains controversial, also because of an uncertain boundary between OA and paywall articles.

Background studies

Literature that endorses the occurrence of the open access citation advantage

A part of the literature endorses the claim that OA publications are prone to show a higher impact, which also reflects in a sort of citation premium (sometimes named as the OA citation advantage, alternatively referred to as the OA citation effect). That premium should be due to the fact that OA documents can easily reach a wider audience in comparison to research findings only accessible behind a paywall, and they are also subject to higher immediate and long-run recognition (Eysenbach 2006; MacCallum and Parthasarathy 2006; Wang et al. 2015).

More than a dozen studies find OA articles to be more cited than their toll-access counterpart focusing on specific disciplines such as computer science (Lawrence 2001), physics (Brody et al. 2004; Harnad et al. 2004), astrophysics (Henneken et al. 2006; Metcalfe 2005, 2006; Schwarz and Kennicutt 2004); and several others (Antelman 2004; Atchison

and Bull 2014; Evans and Reimer 2009; Norris et al. 2008; Riera and Aibar 2013; Sotudeh et al. 2015). Hajjem et al. (2005) argue that, although differences are depending on discipline, the OA citation advantage is often so large that it is unrealistically an outgrowth of a self-selection bias only, namely, that higher-quality articles are made available in OA mode by author choice. Björk and Solomon (2012) find higher citation rates for subscription journals, but they fade away once accounting for confounding factors such as discipline, publishers' country, and journals' age. The study of Greyson et al. (2014) adjusts the analysis for other potential confounders and finds that OA archived articles are more likely to be cited in comparison to non-OA documents.

Due to the interest in the topic, the studies have also branched into several related strands, such as the speculations about the relationship between OA and the Impact Factor (Björk and Solomon 2012; Gumpenberger et al. 2013; Gunasekaran and Arunachalam 2012, 2014; Lin 2009; Mahesh 2012; Pisoschi and Pisoschi 2016; Pollock and Michael 2019; Rordorf 2010) or other metrics (Solomon et al. 2013), the investigations of the OA citation premium for books (Snijder 2010, 2016), the effects of the availability of OA copies through search engines and scholarly social networks (Jamali and Nabavi 2015; Niyazov et al. 2016; Pitol and De Groote 2014), and the building of simulation models based on the OA advantage postulate (Bernius et al. 2013; Bernius and Hanauske 2009).

Literature that raises doubts about the occurrence of the open access citation advantage

Conversely, especially starting from the second half of the 2000s, other authors contend that the relationship between OA and higher citation achievements lacks empirical evidence (Davis 2011; Davis et al. 2008; Dorta-González et al. 2017; Ingwersen and Elleby 2011; Kurtz et al. 2005; Sotudeh and Horri 2007) or that, if any advantage exists, OA is not the primary cause (Calver and Bradley 2009; Craig et al. 2007; Davis 2009; Davis and Fromerth 2007; Gaulé and Maystre 2011; Moed 2007; Zhang and Watson 2017; Zhang 2006). Early skepticism on the advantages consequent to the online, open availability of articles can be found in Anderson et al. (2001). However, it deserves mentioning that the above study refers to an experiment dating back to a bygone era when print-only journals were still predominant.

Wray (2016a) questions the validity of the results presented in a study referenced in the previous section (Sotudeh et al. 2015), arguing that the alleged OA citation advantage in the field of social sciences and humanities is "just an artifact of the methods … employed" (p. 1033). Incidentally, Wray's (2016a) position gave also birth to a discussion on other alleged benefits carried by the adoption of OA (Polonioli 2016; Wray 2016b). Besides, some works establish a relationship between the alleged OA citation advantage and the rank of the publication outlets (Koler-Povh et al. 2014; McCabe and Snyder 2014; Xia and Nakanishi 2012).

Overview of the background studies

The results presented in the above-referenced studies do not converge. Different positions on the topic are passionately discussed, without being able to get to the bottom of the matter. The disagreement on the actual occurrence of the OA citation advantage emerges clearly from the literature reviews published so far (Craig et al. 2007; Davis and Walters 2011; Joint 2009; McKiernan et al. 2016; Swan 2010; Turk 2008; Wagner 2010).

The essential characteristics of the examined works (type and number of articles, source for citation analysis), as well as their main results, have been organized in two tables. The first

Table 1 Studies supporting the occurrence of the OA citation premium

	Type and number of articles [1]				Source for citation analysis [2]									Cit. adv. (pct.) [3]	Results
	GoldJ	GoldA	GreenA	noOA	WoS	ES	GS	SE	RI	ADS	SN	AMS	APS		
Atchison and Bull (2014)		404		323			x							39–1,535%	OA citation advantage is found in political science publishing
Brody, Stamerjohanns, Vallières, Harnad, Gingras, Oppenheim (2004)			95,012	n.a.	x									218–557%	Self-archived articles have up to four times as much impact
Metcalfe [Travis] (2005)			706	476						x				55–245%	Pre-print server boosts citations by a factor of two
Hajjem, Harnad, Gingras (2005)			1,307,038		x									36–172%	Clear and consistent OA citation advantage
Metcalfe [Travis] (2006)			13	157						x				74–167%	Pre-print archiving boosts citations by a factor of 1.7–2.6
Lawrence (2001)			119,924						x					157%	Higher citation rates for papers available online
Pitol and De Groote (2014)			690	292			x							4–144%	Full-text availability significantly increases citations
Schwarz and Kennicutt (2004)			484	311						x				105%	Papers in pre-print server are cited twice as frequently
Norris, Oppenheim, Rowland (2008)	2280			2353	x									44–103%	OA advantage but unknown determinants
Jamali and Nabavi (2015)	4426			2818			x							(25)–97%	Full-text availability in a scholarly social network increases citations
Antelman (2004)			802	1215	x									45–91%	Extent of the citation advantage depends on discipline
Niyazov, Vogel, Price et al. (2016)			5491	25,725			x							16–69%	Paper available in a scholarly social network are cited more
Wang, Liu, Mao, Fang (2015)			587	1174									x	11–52%	OA also benefits from higher attention for a longer period
Eysenbach (2006)	212			1280	x									29–42%	Strong evidence of citation premium
Riera and Aibar (2013)	17	22	122		x									13–38%	OA, along with first author's standing, increases impact
Greyson, Morgan, Hanley, Wahyuni (2014)			358	1565	x									29%	Huge citation advantages for self-archived articles
Evans and Reimer (2009)			26,002,796		x									~0–25%	OA shifts upward citations, less than previously thought
Piwowar, Priem, Larivière, Alperin, et al. (2018)	100,000				x									18%	OA advantage driven by Green and hybrid OA articles
Gargouri et al. (2010)			27,197		x									n.a.	Advantage for both mandated and self-selected OA articles
Xia, Myers, Wilhoite (2011)	486			389	x									n.a.	Multiple OA availability raises citation count

[1] *GoldJ* articles in Gold OA journals; *GoldA* Gold OA articles in hybrid journals; *GreenA* Green OA articles (mandated or optional self-archiving); *noOA* non-OA articles. Note that, in some studies, the type of OA articles is not exactly specified

[2] *WoS* Web of Science; *ES* Elsevier's Scopus; *GS* Google Scholar; *SE* search engines such as Google and Yahoo; *RI* ResearchIndex (also known as CiteSeer); *ADS* NASA's Astrophysics Data System; *SN* Springer Nature online publishing platform; *AMS* American Mathematical Society, MathSciNet; *APS* American Physiological Society, HighWire Press

[3] Open access citation advantage in percentage: results presented in the examined works and/or author's study based on raw data presented there, negative numbers in brackets

includes the studies supporting the existence of the OA citation premium (Table 1), while the second features the studies questioning its occurrence (Table 2). In each table, the referenced works are presented in descending order according to the upper bound of the estimated citation advantage. The following points arise in light of an at-a-glance comparison. Most of the empirical findings—both supporting and questioning the OA citation premium—rely on data collected from Web of Science (WoS), formerly run by ISI Thomson Reuter, now by Clarivate Analytics. Other specialized sources are seldom used, and they typically lead to endorse the existence of a significant OA citation supremacy. When the data sources include Google Scholar (GS), it is not unusual to encounter very large estimates of the OA advantage. Indeed, the largest values in both the tables—a four-digit and three-digit citation advantage, respectively—are the result of investigations conducted based on GS.

Within the growing literature dealing with the OA citation advantage, a few studies point out two related issues: the variety of practices involved in the OA notion is the former, the different audience possibly reached by OA and subscription-based articles is the latter. They are addressed in the following subsections.

Table 2 Studies questioning the occurrence of the OA citation premium

	Type and number of articles [1]				Source for citation analysis [2]									Cit. adv. (pct.) [3]	Results
	GoldJ	GoldA	GreenA	noOA	WoS	ES	GS	SE	RI	ADS	SN	AMS	APS		
Koler-Povh, Juznic, Turk (2014)		442		1584	x	x	x							(100)-518%	OA alone is not sufficient for the citation advantage
Dorta-González, González-Betancor, Dorta-González (2017)		60,566	1,077,826		x									(100)-229%	No OA advantage both at article and journal level
Yanjun Zhang (2006)	49			45	x			x						(17)-78%	OA has more web citations, not from scholarly publications
Moed (2007)			18,757		x									2-67%	Quality bias and early view effect, no citation premium
Gaulé, Maystre (2011)		723		3665	x									33-53%	No evidence for a causal effect of OA on citations
Xia and Nakanishi (2012)		200		467	x									0-44%	OA favors bottom-rank journals, no advantage for top journals
Davis and Fromerth (2007)			511	2254								x		35%	Inferential support for quality gap but not for OA citation advantage
Davis [Philip] (2009)		1,613		9400	x									11-32%	OA advantage is uncertain and decreasing over time
Ottaviani [Jim] (2016)			3850	3850	x									(47)-19%	OA advantage, if any, exists especially for better articles
Davis [Philip] (2011)		712		2533	x									~1-11%	OA articles reach a wider audience but are not more cited
McCabe and Snyder (2014)		> 200,000			x									(8)-8%	Modest clues of OA advantages for top-rank journals
Zhang and Watson (2017)	179		67	404	x									(27)-4%	Comparable citation rates for OA and non-OA articles
Davis, Lewenstein, Simon et al. (2008)		247		1372								x		(5)-(10)%	No proof of citation advantage for OA articles
Kurtz, Eichhorn, Accomazzi, Grant, Demleitner, Henneken, Murray (2005)			2592							x				n.a.	Unneeded OA premium in a specific discipline

[1] *GoldJ* articles in Gold OA journals; *GoldA* Gold OA articles in hybrid journals; *GreenA* Green OA articles (mandated or optional self-archiving); *noOA* non-OA articles. Note that, in some studies, the type of OA articles is not exactly specified

[2] *WoS* Web of Science; *ES* Elsevier's Scopus; *GS* Google Scholar; *SE* search engines such as Google and Yahoo; *RI* ResearchIndex (also known as CiteSeer); *ADS* NASA's Astrophysics Data System; *SN* Springer Nature online publishing platform; *AMS* American Mathematical Society, MathSciNet; *APS* American Physiological Society, HighWire Press

[3] Open access citation advantage in percentage: results presented in the examined works and/or author's study based on raw data presented there, negative numbers in brackets

Related issues of the citation advantage: open access definition

With regard to the first issue, whether OA benefits or not of a citation premium may be a matter of confounding factors and measurement problems. Two pertinent examples are the self-selection bias—namely, author self-selection of high-quality articles to be made openly available—and the early view effect—that is to say, the effect due to early availability of documents, especially in the form of pre-prints. For instance, Moed (2007) finds evidence of an early access effect for the articles deposited in pre-print servers prior to publication and, once correcting for it, concludes that no citation advantage takes place. Gargouri et al. (2010), analyzing more than 27 thousand documents published in nearly two thousand journals over a five-year time span, show that the OA advantage, albeit significant, goes hand in hand with the general, positive citation gap due to mere high quality. Besides finding no evidence of a causal nexus between OA and impact, Gaulé and Maystre (2011) corroborate the theory that links the citation premium to self-selection bias. Craig et al. (2007) stress the fact that self-selection and early view can be jointly as significant as to make the residual OA effect negligible. Ottaviani (2016) takes into account both of them and, although finding in some cases the occurrence of an OA advantage, highlights that the citation premium goes to the benefit of the best pieces.

A further confounding factor consists of the weak definition of what OA really means and where its boundary lies. Piwowar et al. (2018) provide a non-comprehensive list in which eight well-established OA subtypes are enumerated. Gold OA refers to the practice

of paying Article Processing Charge (APC) by the authors or their supporters upon publication so that a document is made immediately available to the largest audience. Gold OA articles can be found both in Gold OA and hybrid journals. Green OA concerns the optional or mandated practice of self-archiving a publication, either the pre-print or post-production version, in pre-print servers or institutional repositories, perhaps after an embargo period. It deserves mentioning that scholarly social networks allow the users to store private full-texts and to share them. Other forms of direct sharing (Gardner and Gardner 2015) and the so-called Black OA phenomenon (Björk 2017; Piwowar et al. 2018)—namely, the activity of illegal download, mainly through Sci-Hub (Bohannon 2016)—are contributing to shaping a de-facto OA landscape. The previous Tables 1 and 2 make it clear that the literature referenced above focuses on different OA subtypes, which may be among the reasons behind the diverging results on the OA citation advantage. For instance, an OA citation advantage is reported in the piece of Piwowar et al. (2018), where it is also pointed out that OA articles in hybrid journals and self-archived copies are the main drivers of that effect. Similarly, Miguel et al. (2011) highlight that nearly the same level of cites per document characterizes Gold OA journals and traditional subscription-based publication outlets, while Green OA benefits from about twice as much citations per article. Instead, Zhang and Watson (2017) show that, when controlling for publication year, comparable citation rates characterize Gold OA, Green OA, and non-OA articles, and it is uncertain that papers published in Gold OA journals accrue more citations than Gold OA documents featured in hybrid journals. Incidentally, it is not even clear whether or not the OA type really matters. Indeed, Kurtz et al. (2005) hypothesize a relationship between their finding of no OA citation advantage and the fact that, in well-funded disciplines, authors are already provided with full access to the relevant literature, regardless of how the same literature is made available.

Related issues of the citation advantage: open access audience

The second issue essentially involves how OA publications are accessed and what kind of audience they reach. According to Davis et al. (2008), although no proof of citation advantage is found, OA articles are more accessed than paywall articles, especially in terms of full-text downloads. Xia et al. (2011) offer an interesting viewpoint: citation count seems to correlate with the "number of web search engines that… return a link to the free full text" (p. 22). However, the authors do not test if the number of accessible copies—namely, the multiple OA availability of a paper, through several distribution channels and platforms—raises the impact. On the other hand, both Pitol and De Groote (2014) and Jamali and Nabavi (2015) find a weak positive correlation between the number of available versions and citations, which is also confirmed in later studies (Martín-Martín et al. 2015). Zhang (2006) compares articles published in an OA journal and a non-OA journal in the field of computer technology-based communication. The former receives nearly twice as much as web citations than the latter. However, the bulk of them does not come from scholarly publication sources. Instead, they are mostly from non-authoritative documents, such as online articles and discussions, unpublished works, and so forth. The same idea mirrors what has been found for the relationship between OA and the open encyclopedia Wikipedia (Teplitskiy et al. 2017), and is also at the root of the conclusions put forth by Davis (2011), who claims that OA articles reach a wider audience but are not more cited in the specialized literature, thus OA represents a benefit for the community at large, not necessarily for the scientific community.

The issue of the reached audience brings with itself a secondary aspect. Although the OA citation advantage is not an artifact related to the indexing database used, the results reported in

previous Tables 1 and 2 suggest that resorting to non-selective search engines—such as GS—may magnify it in comparison to selective (and subscription-based) indexing services—such as WoS and Elsevier's Scopus (ES). GS is known to provide better coverage for sources mostly disregarded by WoS and ES (Meho and Yang 2007; Mingers and Leydesdorff 2015; Waltman 2016). That also reflects in citation counts (Kousha and Thelwall 2007). According to Martín-Martín et al. (2015) and their study of a large sample of highly-cited publications, "91.6% of the documents have received more citations in GS than in WoS…. Furthermore, the average number of citations per document in GS is 1790, and 1080 in WoS, which means that on average, GS has 70% more citations" (p. 33). Jacso (2006) stresses that "The inflated counts [of citations] reported by GS are partly due to the inclusion of non-scholarly sources, like promotional pages, table of contents pages, course reading lists" (p. 300), as well as that "self-archiving and other types of web postings" (p. 300) contribute to that outcome. In a later article, the author highlights that GS is affected by the issue of "duplicate, triplicate and quadruplicate records for the same source documents" (Jacsó 2008, p. 102), which reflects in the phenomenon of "stray citations" (Harzing and Alakangas 2016, p. 802). That phenomenon is also likely to push citation counts up for the documents referenced in the duplicate items. Borrowing the words of Mingers and Leydesdorff (2015), "the coverage of research outputs … is much higher in GS … and … GS has a comparatively greater advantage in the non-science subjects" (p. 3). The same authors point out that "the data quality in GS is very poor with … many of the citations coming from a variety of non-research sources" (p. 3). Several other works express concerns about the quality of data retrieved from GS (Beel and Gipp 2010; de Winter et al. 2014; Delgado López-Cózar et al. 2014; Labbé 2010) and call into question its reliability as a provider of bibliometric data (Aguillo 2012; Bohannon 2014; Jacsó 2010; Waltman 2016). Another study (Bar-Ilan 2008) reports that GS strongly references articles stored in pre-print repositories, which are not necessarily peer-reviewed. In a nutshell, OA articles may better meet the crawling ability of GS (Kousha and Thelwall 2008) in comparison to the "academic invisible web" made of the "articles stored in publishers' databases" (Beel et al. 2010, p. 178).

Method, data collection, and processing

Methodological setting

Here, the primary aim is not to bring new evidence in favor or against the citation premium supposedly characterizing OA articles. Rather, this contribution is meant to test the following hypothesis: OA papers accrue more citations in the open indexing database GS than in the selective indexing engine ES. Building on the empirical finding that GS catches citations more easily since it searches throughout the web, and not only in a given set of literature sources, the assumption to be tested here can also be expressed as follows. OA may actually provide an advantage in terms of higher citations, but those additional citations are unevenly distributed between literature sources (peer-reviewed documents) and other sources (working papers, pre-print versions, lecture notes, patents, and so forth), being preeminent in the second ones.

Let us define s_i the number of citations an article i has received in ES over a set period, and g_i the number of citations the same article i has received in GS over the same time span, such that the ratio r_i is given as follows:

$$r_i = g_i/s_i \tag{1}$$

Since citations in GS are expected to be higher than citations in ES, the ratio in Eq. (1) is likely to be above unity. Let us also distinguish between OA documents and articles

published in subscription-based (SB) journals. The former Eq. (1) can be split and rewritten in the following forms:

$$r_{i'}^{OA} = g_{i'}^{OA}/s_{i'}^{OA} \tag{2}$$

$$r_{i''}^{SB} = g_{i''}^{SB}/s_{i''}^{SB} \tag{3}$$

Thus, the hypothesis to be tested is that the average ratio for OA articles is higher than that of SB articles, and the difference is strong enough to support the existence of some kind of relationship:

$$H_0 : \sum_{i'=1}^{n'} r_{i'}^{OA}/n' \simeq \sum_{i''=1}^{n''} r_{i''}^{SB}/n'' \tag{4}$$

$$H_1 : \sum_{i'=1}^{n'} r_{i'}^{OA}/n' \neq \sum_{i''=1}^{n''} r_{i''}^{SB}/n'' \tag{5}$$

In other words, the ratio of citation counts from GS to citation counts from ES is compared between OA and subscription-based articles, with the null hypothesis H_0 that the two ratios are nearly equal. The Analysis of Variance (ANOVA) fits the purpose of testing whether the alternative hypothesis H_1 of Eq. (5) should be rejected or not.

The distribution of citation counts is known to be highly skewed (Albarrán et al. 2011; Seglen 1992). The same problem is likely to affect this study, not only because of the focus on citation counts but also because the ratios above have a natural lower bound—namely, they cannot fall below zero—while having no definite upper limit. ANOVA relies on an F-test, which is not robust against violations of normality. Although ANOVA is thought to tolerate departures from the normality assumption, the same assumption can be relaxed especially for large samples. Accordingly, following the approach adopted in several referenced studies (Eysenbach 2006; Gargouri et al. 2010), the logarithmic transformation is used to normalize the data, or at least reduce the effect due to outliers in long-tailed distributions:

$$H_0 : \sum_{i'=1}^{n'} \left(\ln r_{i'}^{OA}\right)/n' \simeq \sum_{i''=1}^{n''} \left(\ln r_{i''}^{SB}\right)/n'' \tag{6}$$

$$H_1 : \sum_{i'=1}^{n'} \left(\ln r_{i'}^{OA}\right)/n' \neq \sum_{i''=1}^{n''} \left(\ln r_{i''}^{SB}\right)/n'' \tag{7}$$

In the analysis below, the occurrence of the alternative hypothesis H_1 as in Eq. (7) is tested setting several citation thresholds, namely, varying the minimum numbers of citations featured in ES. That is because the occurrence of the hypothesized relationship is checked using ratios and the number of citations in ES is their denominator. Hence, a low number of cites from Elsevier's database may lead to biased results. Here are a few pertinent examples. Let us suppose first that an article features one citation in ES and two in GS so that $r_i = 2$ according to Eq. (1). Let us then suppose that an article accrues two cites in ES and three in GS. Although the citation difference is the same as in the previous case, Eq. (1) provides a ratio r_i equal to 1.5. Lastly, let us suppose an articles is cited ten and

Table 3 Summary of the case studies

Level	Case study	Time window	Articles[a]	Gold OA[a]		Green OA[a]	
			n.	n.	pct.	n.	pct.
Journal	Scientometrics	2013–2017	1615	129	8.0%	892	55.2%
Topic	Greenhouse gas emissions	2015	332	65	19.6%	121	36.4%
Author	P. Greengard	1985–2018	673	284	42.2%	198	29.4%

[a]Number of articles with at least one citation in Scopus

eleven times in ES and GS, respectively. Again, there is no change in the citation difference between the documents, but the ratio r_i in Eq. (1) takes on the value of 1.1. The relationship tested here is expected to get a stronger confirmation if the alternative hypothesis H_1 of Eq. (7) should not be rejected not only for low citation thresholds but also limiting the analysis to highly-cited documents. Obviously, a side effect has to be considered: raising the threshold, the sample size shrinks and the test power gets worse.

Case studies

The null and alternative hypotheses above are tested on the three following case studies, which can roughly represent the journal level, the topic level, and the author level, respectively (Table 3). The first includes the articles published in the journal *Scientometrics* during the period 2013–2017 and featured in the indexing database ES. The second test is performed on all the articles, published during the year 2015 and indexed in ES, that use the expression "greenhouse gas emissions" in the title. The third case focuses on the ES indexed articles authored by a highly-prolific scholar, the Nobel Prize winner P. Greengard, during an about three-decade time window that goes from 1985 to 2018.

Data gathering process

ES data was gathered through the export function of search results after performing the following queries:

- SRCTITLE(scientometrics) AND PUBYEAR > 2012 AND PUBYEAR < 2018 AND (LIMIT-TO(EXACTSRCTITLE,"Scientometrics"));
- TITLE("greenhouse gas emissions") AND PUBYEAR = 2015;
- AUTHOR-NAME(greengard, AND p) AND PUBYEAR > 1984 AND PUBYEAR < 2019.

The variables of interest are "Titles", "Cited by", namely, the number of citations recorded in the indexing database, and "Access type", which has been used in order to identify Gold OA articles.

GS data was collected using the software Publish or Perish (Harzing 2007; Harzing and van der Wal 2008) performing the following Google Scholar queries:

- Years: 2017–2017 (then Years: 2016–2016; 2015–2015; 2014–2014; 2013–2013) AND Publication/Journal: Scientometrics (each year was searched separately to bypass the limit of 1000 results per search);

- Years: 2015–2015 AND Title words: greenhouse gas emissions;
- Authors: Greengard Paul.

The variables of interest are "Cites", thus the number of citations accrued by each article according to GS, and "Titles" and "ArticleURL", which have been used in succession to search for Green OA versions, partly following the search strategy adopted in the previously referenced literature (Xia and Nakanishi 2012). As far as the latter point is concerned, the title of non-Gold OA documents was copied/pasted into the GS search box. Once the right-hand frame of the result page had provided a link to a self-archived PDF or HTML version, the link was clicked through to check whether the full-text copy was actually available and accessible. Otherwise, the URL of the article was used to get access to its page on the publisher's website and verify whether the Unpaywall extension, developed by H. Piwowar and J. Priem (Else 2018a), could locate an OA copy. Once the Unpaywall icon's color had turned to green, it was clicked through to verify the actual accessibility of the full-text copy.

The articles extracted from ES and GS has been matched in a worksheet using the titles. Exact matches have been found for 80% (case study: *Scientometrics*), 58% (case study: greenhouse gas emission), and 52% (case study: P. Greengard) of the observations. Mismatches were essentially due to the different way in which ES and GS record the symbols (e.g., Greek letters) and the punctuation marks (e.g., commas, colons, hyphens with or without spaces before and after, and so forth) used in the titles, and they have been corrected manually. Except for three documents published in *Scientometrics* and eight authored by P. Greengard, all the articles extracted from ES was found also in GS. Data collection was performed between March and April 2019.

A caveat is in order here, hence, let me make a brief digression. Both the sources of bibliometric information are known to be characterized by several drawbacks and short-comings. On the one hand, duplicate items (Valderrama-Zurián et al. 2015) and matching issues (Franceschini et al. 2016a, b; Meester et al. 2016) affect ES. On the other hand, duplicate items (Jacsó 2008) as well as "content gaps, incorrect citation counts, and phantom data" (Waltman 2016, p. 269) are among the topical issues listed in the literature with regard to GS. As far as the analysis below is concerned, the problem of duplicate items has been observed, especially as regards GS. Although the mismatches have been corrected when possible, the findings should be taken with due caution.

Another caveat refers to the representativeness of the sample. Two matters of concerns lie in the fact that it is not randomly chosen and it has a limited size. With regard to the latter aspect, automated searches of Green OA articles return rough results. In order to achieve accurate classifications, those searches are better performed manually using the tools mentioned above—GS and Unpaywall—and checking the reliability of each result. Nevertheless, inspecting the web in search for self-archived copies is a laborious, time-consuming process (Else 2018b), which hinders the ability to increase the sample size and also negatively affects the generalizability of the results.

Results

A focus on Gold open access

In the first instance, are considered as OA only the articles published using the so-called author-pays model in Gold OA journals as well as in hybrid publication outlets. On the opposite side stand all the remaining articles, including those available as Green OA.

The detailed results are presented in "Appendix 1"—Tables 4, 5 and 6 for Eqs. (6) and (7)—and are quite disappointing. Regardless of the citation threshold, the average GS-to-ES citation ratio for Gold OA articles—namely, the left-hand term in Eqs. (6) and (7)—is always higher than the mean GS-to-ES citation ratio for non-OA items—i.e., the right-hand term in Eqs. (6) and (7)—which would confirm the hypothesis under scrutiny. For instance, in the case of *Scientometrics*, the ratio for Gold OA articles varies between 0.503 (citation threshold = 5) and 0.587 (citation threshold = 1). It means that, on average, the number of citations to Gold OA papers featured in GS is near twice the cites one can find in ES. Meanwhile, the same ratio for non-OA documents ranges between 0.478 (citation threshold = 10) and 0.523 (citation threshold = 1).

Nonetheless, the difference is seldom significant. That is particularly evident in the case study focusing on the 2015 articles that include the wording "greenhouse gas emissions" in the title. On the contrary, a notable exception to the generally weak statistical significance of the results is represented by the case study of the works authored by P. Greengard. As a case in point, considering a minimum citation number in ES equal to three, the GS-to-ES citation ratio for OA articles is 0.270 with a variance of 0.016, while the same rate for non-OA documents is 0.219 with a variance of 0.126, and the difference is significant (F-stat 5.118, p value 0.0240). Comparable results are found for citations thresholds from two to five.

The role played by Green open access

The analysis discussed in the previous section adopts a very stringent definition of OA. That may be excessively restrictive since OA manifests itself following several routes, shaping a de facto OA landscape behind and beyond its known market share (Wren 2005). Accordingly, below I present a variant of the analysis, where the OA status is determined by searching in the web whether a full-text is made available by the author(s) or not. I refer to any kind of full-text: pre-prints or post-prints, deposited in an institutional repository or circulated through a scholarly social network, and so forth.

The detailed results can be found in Appendix 2—Tables 7, 8 and 9 for Eqs. (6) and (7)—and are quite more satisfactory than those previously presented. In the case study of the journal *Scientometrics*, the GS-to-ES citation ratio is significantly different when comparing OA and non-OA articles for citation thresholds up to eleven. As a case in point, let us consider a minimum number of citations in ES equal to seven. The GS-to-ES citation ratio for OA articles is 0.498 with a variance of 0.066, while the rate for non-OA documents is 0.443 with a deviation of 0.056. Although that may seem a small difference, it is significant (F-stat 8.041, p value 0.0047). A look at the underlying data may help in understanding why the alternative hypothesis H_1 of Eq. (7) is not rejected. For the same citation threshold of seven, the GS-to-ES citation ratio for OA papers takes on a median value of 0.49 and it varies up to a maximum of 1.77 (Fig. 1). Instead, the same rate for non-OA articles shows a median value of 0.44 but its maximum amount does not overtake 0.97. It means that OA documents get up to six times more citations in GS in comparison to ES, while their non-OA counterparts are cited in GS only three times more than in ES.

Comparable results are found for the case study focusing on the articles published in 2015 that include the expression "greenhouse gas emissions" in the title (Fig. 2). For low citation thresholds, especially below eight, the difference between the left-hand and the right-hand terms of Eq. (7) is significant ($0.0000 \leq p$ value ≤ 0.0064). Instead, in the case study focusing on the works authored by P. Greengard, the statistical significance of the results is worse than that discussed in the previous section.

Fig. 1 Mean values and ranges of the GS-to-ES citation ratios for articles published in *Scientometrics* (time windows: 2013–2017), natural logarithms of the ratios
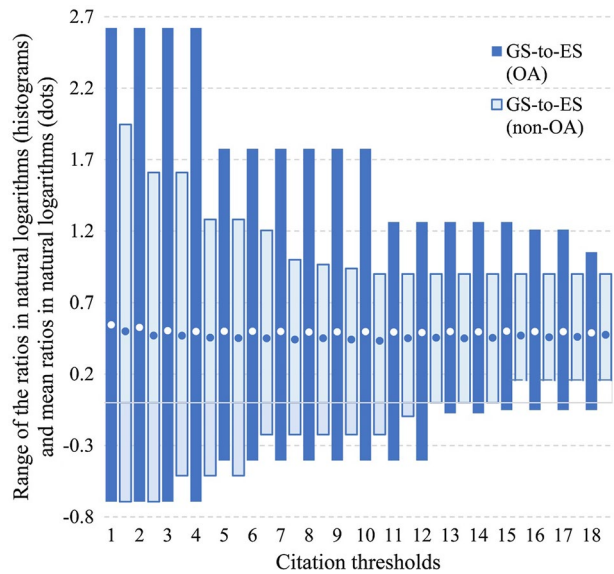

Fig. 2 Mean values and ranges of the GS-to-ES citation ratios for articles that include the expression "greenhouse gas emissions" in the title (time window: 2015), natural logarithms of the ratios
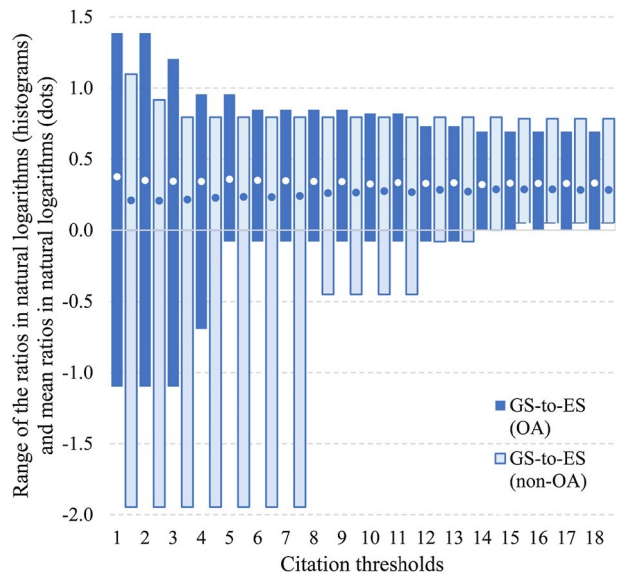
Conclusions

The move towards OA has engendered a core of genuine issues still far from being solved. They range from the ethical side—namely, the legitimacy of some players that are taking part in the game and playing a questionable role in the publishing industry (Beall 2012)—up to the financial viewpoint—i.e., the long-term viability of the business models in scientific publishing, and their suitability to the budget constraints

experienced by scholarly libraries, with non-negligible quality-side concerns (Jeon and Rochet 2010; McCabe and Snyder 2005, 2007, 2012). The multifaceted literature growingly addressing all those issues seldom agrees on the origins of the problems, let alone the solutions.

Here I focus only on a narrow range of issues, which are essentially hinged on the citation behavior of OA article users. On the one hand, this study builds on the citation premium that is supposed to characterize OA publications in comparison to the documents accessible behind a paywall only. On the other hand, it relies on the fact that an open indexing service such as GS tends to catch more citations than traditional selective indexing databases like ES. The hypothesis under scrutiny is that the higher level of citations accrued by OA papers in comparison to their non-OA counterparts may be unevenly distributed between different kinds of sources. Put differently, the density of citations received by OA documents may be higher in non-peer-reviewed literature such as working papers, lecture notes, blog pieces, and so forth, precisely those kinds of sources that easily fit the crawling ability of GS.

The results presented in the previous section suggest that the hypothesized relationship is neither so strong as to be generally accepted nor so weak as to be neglected. Although some OA documents benefit from a citation premium that may be emphasized by the openness and inclusiveness of the indexing database, the main limitations are as follows. The relationship takes place in two out of the three analyzed case studies, while it is essentially absent in the other case study. Statistically significant evidence can be achieved only by setting low to medium citation thresholds, thus the relationship fades away when looking at the documents that accrue many cites. The relationship cannot be ascribed with certainty to how the indexing databases work rather than to their drawbacks, such as duplicate items and incorrect citation counts.

Nonetheless, this study also highlights two aspects that deserve further consideration. The first concerns the role played by GS as a citation source in comparison to ES. The former catches citations more easily than the latter but data quality remains an open issue. While the citations counted in ES indicate recognition in the scientific literature, it is unclear the scope of citations recorded by GS and the kind of impact they are meant to represent. The second aspect refers to the distinction between Gold and Green OA, which affects the results achieved here. While focusing on the Gold OA route leads to inconclusive evidence, the joint consideration of Gold and Green OA provides some confirmation of the tested hypothesis. Whether the assumed relationship—between the OA citation premium and the openness and inclusiveness of GS—takes place or not also depends on where the OA boundary lies, as well as on how articles are clustered into the OA subtypes.

Appendix 1: Results for Gold open access

See Tables 4, 5 and 6.

Table 4 Gold OA articles published in Scientometrics (time window: 2013–2017), natural logarithms of the ratios

Citation threshold[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GS-to-ES citation ratio for OA articles																		
n.	126	117	109	93	84	80	74	65	55	44	37	33	33	33	30	27	26	24
Mean	0.587	0.555	0.545	0.522	0.503	0.511	0.505	0.506	0.511	0.515	0.525	0.506	0.506	0.506	0.529	0.546	0.553	0.515
Var.	0.147	0.122	0.116	0.110	0.065	0.064	0.060	0.057	0.059	0.061	0.063	0.062	0.062	0.062	0.061	0.062	0.063	0.045
GS-to-ES citation ratio for non-OA articles																		
n.	1480	1340	1182	1025	917	806	725	633	570	494	443	377	334	301	261	236	217	198
Mean	0.523	0.503	0.488	0.481	0.484	0.483	0.480	0.480	0.479	0.478	0.480	0.480	0.484	0.484	0.489	0.482	0.481	0.482
Var.	0.146	0.120	0.097	0.080	0.073	0.069	0.064	0.058	0.058	0.056	0.051	0.052	0.046	0.046	0.039	0.036	0.035	0.035
AnoVa																		
F-stat	3.241	2.429	3.335	1.674	0.410	0.849	0.650	0.702	0.935	1.011	1.313	0.379	0.292	0.316	1.005	2.593	3.216	0.659
p val.	0.072	0.119	0.068	0.196	0.522	0.357	0.420	0.402	0.334	0.315	0.252	0.538	0.589	0.574	0.317	0.109	0.074	0.418
Sign.[2]	*		*														*	

[1] Minimum number of citations received in Scopus

[2] Significance levels: *10%; **5%; ***1%

Table 5 Gold OA articles that include the expression "greenhouse gas emissions" in the title (time window: 2015), natural logarithms of the ratios

Citation threshold[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GS-to-ES citation ratio for OA articles																		
n.	63	57	52	45	41	36	35	33	30	30	28	27	23	23	23	23	22	17
Mean	0.363	0.345	0.329	0.312	0.354	0.355	0.348	0.336	0.340	0.340	0.352	0.351	0.367	0.367	0.367	0.367	0.383	0.388
Var.	0.110	0.084	0.064	0.067	0.042	0.035	0.034	0.028	0.027	0.027	0.025	0.026	0.020	0.020	0.020	0.020	0.014	0.016
GS-to-ES citation ratio for non-OA articles																		
n.	261	236	223	210	197	182	172	156	146	131	122	113	102	88	82	78	67	62
Mean	0.268	0.247	0.251	0.258	0.263	0.252	0.253	0.261	0.258	0.249	0.245	0.248	0.236	0.226	0.227	0.223	0.232	0.230
Var.	0.187	0.177	0.165	0.155	0.162	0.166	0.169	0.150	0.157	0.166	0.176	0.180	0.197	0.216	0.227	0.236	0.216	0.233
AnoVa																		
F-stat	2.633	2.779	1.772	0.762	2.003	2.194	1.764	1.180	1.236	1.415	1.768	1.543	1.941	2.055	1.926	1.956	2.266	1.768
p val.	0.106	0.097	0.184	0.383	0.158	0.140	0.186	0.279	0.268	0.236	0.186	0.216	0.166	0.155	0.168	0.165	0.136	0.188
Sign.[2]		*																

[1] Minimum number of citations received in Scopus

[2] Significance levels: *10%; **5%; ***1%

Table 6 Gold OA articles authored by P. Greengard (time window: 1985–2017), natural logarithms of the ratios

Citation threshold[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GS-to-ES citation ratio for OA articles																		
n.	281	275	275	272	268	266	265	260	257	256	255	254	254	251	248	246	246	242
Mean	0.275	0.270	0.270	0.267	0.265	0.258	0.259	0.257	0.256	0.256	0.255	0.255	0.255	0.256	0.255	0.257	0.257	0.257
Var.	0.026	0.016	0.016	0.016	0.015	0.010	0.010	0.009	0.009	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
GS-to-ES citation ratio for non-OA articles																		
n.	383	375	372	371	367	366	364	360	356	353	348	343	338	329	326	321	317	310
Mean	0.231	0.219	0.219	0.218	0.220	0.219	0.218	0.218	0.219	0.219	0.220	0.220	0.220	0.228	0.227	0.227	0.226	0.228
Var.	0.137	0.126	0.126	0.126	0.125	0.125	0.125	0.125	0.126	0.127	0.128	0.130	0.131	0.125	0.126	0.128	0.129	0.131
AnoVa																		
F-stat	3.453	5.108	5.118	4.851	3.972	3.046	3.396	3.019	2.721	2.638	2.365	2.385	2.379	1.559	1.542	1.703	1.786	1.565
p val.	0.064	0.024	0.024	0.028	0.047	0.081	0.066	0.083	0.100	0.105	0.125	0.123	0.123	0.212	0.215	0.192	0.182	0.211
Sign.[2]	*	**	**	**	**	*	*	*	*									

[1]Minimum number of citations received in Scopus

[2]Significance levels: *10%; **5%; ***1%

Appendix 2: Results for Gold and Green open access

See Tables 7, 8 and 9.

Table 7 Gold OA and Green OA articles published in Scientometrics (time window: 2013–2017), natural logarithms of the ratios

Citation threshold[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GS-to-ES citation ratio for OA articles																		
n.	1015	951	861	762	694	621	570	505	456	402	358	310	279	255	223	200	185	170
Mean	0.545	0.527	0.504	0.499	0.500	0.500	0.498	0.494	0.496	0.497	0.495	0.491	0.498	0.496	0.500	0.498	0.497	0.489
Var.	0.141	0.119	0.097	0.085	0.072	0.071	0.066	0.060	0.060	0.059	0.056	0.056	0.048	0.048	0.041	0.039	0.038	0.036
GS-to-ES citation ratio for non-OA articles																		
n.	592	507	431	357	307	265	229	193	169	136	122	100	88	79	68	63	58	52
Mean	0.500	0.470	0.470	0.456	0.452	0.451	0.443	0.452	0.442	0.433	0.452	0.455	0.451	0.454	0.470	0.458	0.461	0.476
Var.	0.155	0.122	0.102	0.076	0.072	0.063	0.056	0.050	0.049	0.046	0.042	0.043	0.043	0.045	0.042	0.039	0.037	0.037
AnoVa																		
F-stat	5.211	8.824	3.500	5.472	6.881	6.644	8.041	4.392	6.156	7.421	3.242	1.872	3.134	2.222	1.152	1.951	1.528	0.176
p val.	0.023	0.003	0.062	0.019	0.009	0.010	0.005	0.036	0.013	0.007	0.072	0.172	0.077	0.137	0.284	0.164	0.218	0.675
Sign.[2]	**	***	*	**	***	**	***	**	**	***	*		*					

[1] Minimum number of citations received in Scopus

[2] Significance levels: *10%; **5%; ***1%

Scientometrics (2019) 121:995–1018

Table 8 Gold OA and Green OA articles that include the expression "greenhouse gas emissions" in the title (time window: 2015), natural logarithms of the ratios

Citation threshold[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GS-to-ES citation ratio for OA articles																		
n.	183	161	151	139	130	116	109	101	91	87	79	76	69	61	58	57	52	44
Mean	0.376	0.350	0.344	0.342	0.358	0.350	0.347	0.343	0.341	0.325	0.335	0.329	0.333	0.320	0.330	0.329	0.329	0.331
Var.	0.118	0.093	0.074	0.057	0.049	0.041	0.042	0.040	0.041	0.036	0.035	0.033	0.033	0.026	0.026	0.026	0.025	0.027
GS-to-ES citation ratio for non-OA articles																		
n.	142	132	124	116	108	102	98	88	85	74	71	64	56	50	47	44	37	35
Mean	0.210	0.207	0.216	0.228	0.235	0.233	0.241	0.261	0.265	0.275	0.268	0.284	0.272	0.287	0.287	0.288	0.283	0.283
Var.	0.112	0.109	0.095	0.093	0.094	0.097	0.092	0.042	0.041	0.043	0.043	0.035	0.037	0.037	0.031	0.030	0.033	0.034
AnoVa																		
F-stat	19.00	14.78	13.52	11.23	12.92	11.08	8.897	7.615	6.284	2.562	4.365	2.060	3.321	0.970	1.707	1.543	1.583	1.450
p val.	0.000	0.000	0.000	0.001	0.000	0.001	0.003	0.006	0.013	0.111	0.038	0.153	0.071	0.327	0.194	0.217	0.212	0.232
Sign.[2]	***	***	***	***	***	***	***	***	**		**		*					

[1]Minimum number of citations received in Scopus

[2]Significance levels: *10%; **5%; ***1%

Table 9 Gold OA and Green OA articles authored by P. Greengard (time window: 1985–2017), natural logarithms of the ratios

Citation threshold[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GS-to-ES citation ratio for OA articles																		
n.	476	466	466	463	458	456	455	450	444	442	439	437	434	427	423	420	417	411
Mean	0.260	0.252	0.252	0.250	0.249	0.245	0.245	0.244	0.244	0.243	0.243	0.244	0.243	0.244	0.244	0.245	0.245	0.245
Var.	0.070	0.061	0.061	0.061	0.061	0.058	0.058	0.058	0.058	0.059	0.059	0.059	0.059	0.060	0.060	0.060	0.061	0.061
GS-to-ES citation ratio for non-OA articles																		
n.	188	184	181	180	177	176	174	170	169	167	164	160	158	153	151	147	146	141
Mean	0.224	0.212	0.212	0.209	0.213	0.212	0.210	0.209	0.211	0.212	0.213	0.211	0.214	0.228	0.226	0.225	0.223	0.227
Var.	0.144	0.128	0.128	0.128	0.125	0.126	0.127	0.126	0.126	0.127	0.129	0.132	0.132	0.117	0.118	0.121	0.121	0.124
AnoVa																		
F-stat	2.015	2.582	2.597	2.733	2.028	1.732	2.090	1.988	1.749	1.550	1.445	1.550	1.189	0.419	0.492	0.575	0.700	0.441
p val.	0.156	0.109	0.108	0.099	0.155	0.189	0.149	0.159	0.187	0.214	0.230	0.214	0.276	0.518	0.483	0.449	0.403	0.507
Sign.[2]				*														

[1] Minimum number of citations received in Scopus

[2] Significance levels: *10%; **5%; ***1%

References

Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics, 91*(2), 343–351. https://doi.org/10.1007/s11192-011-0582-8.

Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 subfields and a number of aggregates. *Scientometrics, 88*(2), 385–397. https://doi.org/10.1007/s11192-011-0407-9.

Anderson, K., Sack, J., Krauss, L., & O'Keefe, L. (2001). Publishing online-only peer-reviewed biomedical literature: Three years of citation, author perception, and usage experience. *The Journal of Electronic Publishing*. https://doi.org/10.3998/3336451.0006.303.

Antelman, K. (2004). Do open-access articles have a greater research impact? *College and Research Libraries, 65*(5), 372–382. https://doi.org/10.5860/crl.65.5.372.

Atchison, A., & Bull, J. (2014). Will open access get me cited? An analysis of the efficacy of open access publishing in political science. *PS: Political Science and Politics, 48*(1), 129–137. https://doi.org/10.1017/S1049096514001668.

Bar-Ilan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics, 74*(2), 257–271. https://doi.org/10.1007/s11192-008-0216-y.

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature, 489*(7415), 179. https://doi.org/10.1038/489179a.

Beel, J., & Gipp, B. (2010). Academic search engine spam and Google Scholar's resilience against it. *The Journal of Electronic Publishing*. https://doi.org/10.3998/3336451.0013.305.

Beel, J., Gipp, B., & Wilde, E. (2010). Academic search engine optimization (ASEO). *Journal of Scholarly Publishing, 41*(2), 176–190. https://doi.org/10.3138/jsp.41.2.176.

Bernius, S., & Hanauske, M. (2009). Open access to scientific literature: Increasing citations as an incentive for authors to make their publications freely accessible. In *2009 42nd Hawaii international conference on system sciences* (pp. 1–9). IEEE. https://doi.org/10.1109/hicss.2009.335.

Bernius, S., Hanauske, M., Dugall, B., & König, W. (2013). Exploring the effects of a transition to open access: Insights from a simulation study. *Journal of the American Society for Information Science and Technology, 64*(4), 701–726. https://doi.org/10.1002/asi.22772.

Bernius, S., Hanauske, M., König, W., & Dugall, B. (2009). Open access models and their implications for the players on the scientific publishing market. *Economic Analysis and Policy, 39*(1), 103–116. https://doi.org/10.1016/S0313-5926(09)50046-X.

Björk, B.-C. (2017). Gold, green, and black open access. *Learned Publishing, 30*(2), 173–175. https://doi.org/10.1002/leap.1096.

Björk, B.-C., & Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine, 10*(1), 73. https://doi.org/10.1186/1741-7015-10-73.

Bohannon, J. (2014). Google scholar wins raves: But can it be trusted? *Science, 343*(6166), 14. https://doi.org/10.1126/science.343.6166.14.

Bohannon, J. (2016). Who's downloading pirated papers? Everyone. *Science, 352*(6285), 508–512. https://doi.org/10.1126/science.352.6285.508.

Brody, T., Stamerjohanns, H., Valliéres, F., Harnad, S., Gingras, Y., & Oppenheim, C. (2004). The effect of open access on citation impact. In *National policies on open access (OA) provision for university research output: An international meeting, Southampton*. Southampton. https://doi.org/10.1045/june2004-harnad.

Calver, M. C., & Bradley, J. S. (2009). Patterns of citations of open access and non-open access conservation biology journal papers and book chapters. *Conservation Biology, 24*(3), 872–880. https://doi.org/10.1111/j.1523-1739.2010.01509.x.

Chi Chang, C. (2006). Business models for open access journals publishing. *Online Information Review, 30*(6), 699–713.

Craig, I., Plume, A., McVeigh, M., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics, 1*(3), 239–248. https://doi.org/10.1016/j.joi.2007.04.001.

Davis, P. M. (2009). Author-choice open-access publishing in the biological and medical literature: A citation analysis. *Journal of the American Society for Information Science and Technology, 60*(1), 3–8. https://doi.org/10.1002/asi.20965.

Davis, P. M. (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. *The FASEB Journal, 25*(7), 2129–2134. https://doi.org/10.1096/fj.11-183988.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics, 71*(2), 203–215. https://doi.org/10.1007/s11192-007-1661-8.

Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ, 337*(jul31 1), a568–a568. https://doi.org/10.1136/bmj.a568.

Davis, P. M., & Walters, W. H. (2011). The impact of free access to the scientific literature: A review of recent research. *Journal of the Medical Library Association, 99*(3), 208–217. https://doi.org/10.3163/1536-5050.99.3.008.

de Winter, J. C. F., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of science: A longitudinal study. *Scientometrics, 98*(2), 1547–1565. https://doi.org/10.1007/s11192-013-1089-2.

Delgado López-Cózar, E., Robinson-García, N., & Torres-Salinas, D. (2014). The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology, 65*(3), 446–454. https://doi.org/10.1002/asi.23056.

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, 112*(2), 877–901. https://doi.org/10.1007/s11192-017-2422-y.

Else, H. (2018a). How Unpaywall is transforming open science. *Nature, 560*(7718), 290–291. https://doi.org/10.1038/d41586-018-05968-3.

Else, H. (2018b). How I scraped data from Google Scholar. *Nature*. https://doi.org/10.1038/d41586-018-04190-5.

Evans, J. A., & Reimer, J. (2009). Open Access and global participation in science. *Science, 323*(5917), 1025. https://doi.org/10.1126/science.1154562.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology, 4*(5), e157. https://doi.org/10.1371/journal.pbio.0040157.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016a). The museum of errors/horrors in Scopus. *Journal of Informetrics, 10*(1), 174–182. https://doi.org/10.1016/j.joi.2015.11.006.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016b). Empirical analysis and classification of database errors in Scopus and Web of Science. *Journal of Informetrics, 10*(4), 933–953. https://doi.org/10.1016/j.joi.2016.07.003.

Gardner, C. C., & Gardner, G. J. (2015). Bypassing interlibrary loan via Twitter: An exploration of#icanhazpdf requests. In D. M. Mueller (Ed.), *Acrl 2015 conference proceedings*. Chicago: American Library Association, Association of College and Research Libraries. http://hdl.handle.net/10760/24847. Accessed 17 April 2019

Gargouri, Y., Hajjem, C., Larivière, 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, 5*(10), e13636. https://doi.org/10.1371/journal.pone.0013636.

Gaulé, P., & Maystre, N. (2011). Getting cited: Does open access help? *Research Policy, 40*(10), 1332–1338. https://doi.org/10.1016/j.respol.2011.05.025.

Greyson, D., Morgan, S., Hanley, G., & Wahyuni, D. (2014). Open access archiving and article citations within health services and policy research. *Journal of the Canadian Health Libraries Association/Journal de l'Association des bibliothèques de la santé du Canada*. https://doi.org/10.5596/c09-014.

Gumpenberger, C., Ovalle-Perandones, M.-A., & Gorraiz, J. (2013). On the impact of Gold open access journals. *Scientometrics, 96*(1), 221–238. https://doi.org/10.1007/s11192-012-0902-7.

Gunasekaran, S., & Arunachalam, S. (2012). Impact factors of Indian open access journals rising. *Current Science, 103*(7), 757–760.

Gunasekaran, S., & Arunachalam, S. (2014). The impact factors of open access and subscription journals across fields. *Current Science, 107*(3), 380–388.

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. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 28(4), 39–46. http://arxiv.org/abs/cs/0606079.

Harnad, S., Brody, T., Valliéres, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review, 30*(4), 310–314. https://doi.org/10.1080/00987913.2004.10764930.

Harzing, A.-W. (2007). Publish or Perish. https://harzing.com/resources/publish-or-perish. Accessed April 14, 2019.

Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of science: A longitudinal and cross-disciplinary comparison. *Scientometrics, 106*(2), 787–804. https://doi.org/10.1007/s11192-015-1798-9.

Harzing, A.-W., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics, 8*(1), 61–73. https://doi.org/10.3354/esep00076.

Haug, C. J. (2019). No free lunch: What price plan s for scientific publishing? *New England Journal of Medicine, 380*(12), 1181–1185. https://doi.org/10.1056/NEJMms1900864.

Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Thompson, D., et al. (2006). Effect of E-printing on citation rates in astronomy and physics. *The Journal of Electronic Publishing*. https://doi.org/10.3998/3336451.0009.202.

Ingwersen, P., & Elleby, A. (2011). Do open access working papers attract more citations compared to printed journal articles from the same research unit? In *Proceedings of ISSI 2011: The 13th conference of the international society for scientometrics and informetrics* (Vols. 1 and 2, pp. 327–332).

Jacso, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review, 30*(3), 297–309. https://doi.org/10.1108/14684520610675816.

Jacsó, P. (2008). Google Scholar revisited. *Online Information Review, 32*(1), 102–114. https://doi.org/10.1108/14684520810866010.

Jacsó, P. (2010). Metadata mega mess in Google Scholar. *Online Information Review, 34*(1), 175–191. https://doi.org/10.1108/14684521011024191.

Jamali, H. R., & Nabavi, M. (2015). Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics, 105*(3), 1635–1651. https://doi.org/10.1007/s11192-015-1642-2.

Jeon, D.-S., & Rochet, J.-C. (2010). The pricing of academic journals: A two-sided market perspective. *American Economic Journal: Microeconomics, 2*(2), 222–255. https://doi.org/10.1257/mic.2.2.222.

Joint, N. (2009). The Antaeus column: Does the "open access" advantage exist? A librarian's perspective. *Library Review, 58*(7), 477–481. https://doi.org/10.1108/00242530910978172.

Koler-Povh, T., Južnič, P., & Turk, G. (2014). Impact of open access on citation of scholarly publications in the field of civil engineering. *Scientometrics, 98*(2), 1033–1045. https://doi.org/10.1007/s11192-013-1101-x.

Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology, 58*(7), 1055–1065. https://doi.org/10.1002/asi.20584.

Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics, 74*(2), 273–294. https://doi.org/10.1007/s11192-008-0217-x.

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, 41*(6), 1395–1402. https://doi.org/10.1016/j.ipm.2005.03.010.

Laakso, M., & Björk, B.-C. (2012). Anatomy of open access publishing: A study of longitudinal development and internal structure. *BMC Medicine, 10*(1), 124. https://doi.org/10.1186/1741-7015-10-124.

Laakso, M., & Björk, B.-C. (2016). Hybrid open access: A longitudinal study. *Journal of Informetrics, 10*(4), 919–932. https://doi.org/10.1016/j.joi.2016.08.002.

Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B.-C., & Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PLoS ONE, 6*(6), e20961. https://doi.org/10.1371/journal.pone.0020961.

Labbé, C. (2010). Ike Antkare one of the great stars in the scientific firmament. *ISSI Newsletter, 6*(2), 48–52.

Lawrence, S. (2001). Online or invisible? *Nature, 411*(6837), 521.

Lin, S.-K. (2009). Full open access journals have increased impact factors. *Molecules, 14*(6), 2254–2255. https://doi.org/10.3390/molecules14062254.

MacCallum, C. J., & Parthasarathy, H. (2006). Open access increases citation rate. *PLoS Biology, 4*(5), e176. https://doi.org/10.1371/journal.pbio.0040176.

Mahesh, G. (2012). Open access and impact factors. *Current Science, 103*(6), 610.

Martín-Martín, A., Orduna-Malea, E., Ayllon, J. M., & Delgado López-Cózar, E. (2015). *Dooes Google Scholar contain all highly cited documents (1950–2013)? Granada: EC3 working papers*.

McCabe, M. J., & Snyder, C. M. (2005). Open access and academic journal quality. *American Economic Review, 95*(2), 453–458. https://doi.org/10.1257/000282805774670112.

McCabe, M. J., & Snyder, C. M. (2007). Academic journal prices in a digital age: A two-sided market model academic journal prices in a digital age. *The B.E. Journal of Economic Analysis & Policy, 7*(1), Article 2.

McCabe, M. J., & Snyder, C. M. (2012). The economics of open-access journals. In *10th Annual international industrial organization conference* (p. paper id 34). Arlington: George Mason School of Law.

McCabe, M. J., & Snyder, C. M. (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry, 52*(4), 1284–1300. https://doi.org/10.1111/ecin.12064.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., et al. (2016). How open science helps researchers succeed. *eLife, 5,* 1–19. https://doi.org/10.7554/elife.16800.

Meester, W. J. N., Colledge, L., & Dyas, E. E. (2016). A response to "The museum of errors/horrors in Scopus" by Franceschini et al. *Journal of Informetrics, 10*(2), 569–570. https://doi.org/10.1016/j.joi.2016.04.011.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and Google Scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105–2125. https://doi.org/10.1002/asi.20677.

Metcalfe, T. S. (2005). The rise and citation impact of astro-ph in major journals. *Bulletin of the American Astronomical Society, 37,* 555–557.

Metcalfe, T. S. (2006). The citation impact of digital preprint archives for solar physics papers. *Solar Physics, 239*(1–2), 549–553. https://doi.org/10.1007/s11207-006-0262-7.

Miguel, S., Chinchilla-Rodríguez, Z., & de Moya-Anegón, F. (2011). Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology, 62*(6), 1130–1145. https://doi.org/10.1002/asi.21532.

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research, 246*(1), 1–19. https://doi.org/10.1016/j.ejor.2015.04.002.

Moed, H. F. (2007). The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology, 58*(13), 2047–2054. https://doi.org/10.1002/asi.20663.

Niyazov, Y., Vogel, C., Price, R., Lund, B., Judd, D., Akil, A., et al. (2016). Open access meets discoverability: Citations to articles posted to academia.edu. *PLOS ONE, 11*(2), e0148257. https://doi.org/10.1371/journal.pone.0148257.

Norris, M., Oppenheim, C., & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology, 59*(12), 1963–1972. https://doi.org/10.1002/asi.20898.

Odlyzko, A. (2002). The rapid evolution of scholarly communication. *Learned Publishing, 15*(1), 7–19. https://doi.org/10.1087/095315102753303634.

Ottaviani, J. (2016). The post-embargo open access citation advantage: It exists (probably), it's modest (usually), and the rich get richer (of course). *PLoS ONE, 11*(8), e0159614. https://doi.org/10.1371/journal.pone.0159614.

Pisoschi, A. M., & Pisoschi, C. G. (2016). Is open access the solution to increase the impact of scientific journals? *Scientometrics, 109*(2), 1075–1095. https://doi.org/10.1007/s11192-016-2088-x.

Pitol, S. P., & De Groote, S. L. (2014). Google Scholar versions: Do more versions of an article mean greater impact? *Library Hi Tech, 32*(4), 594–611. https://doi.org/10.1108/LHT-05-2014-0039.

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., et al. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ, 6,* e4375. https://doi.org/10.7717/peerj.4375.

Pollock, D., & Michael, A. (2019). Open access mythbusting: Testing two prevailing assumptions about the effects of open access adoption. *Learned Publishing, 32*(1), 7–12. https://doi.org/10.1002/leap.1209.

Polonioli, A. (2016). Debunking unwarranted defenses of the status quo in the humanities and social sciences. *Scientometrics, 107*(3), 1519–1522. https://doi.org/10.1007/s11192-016-1906-5.

Riera, M., & Aibar, E. (2013). ¿Favorece la publicación en abierto el impacto de los artículos científicos? Un estudio empírico en el ámbito de la medicina intensiva. *Medicina Intensiva, 37*(4), 232–240. https://doi.org/10.1016/j.medin.2012.04.002.

Rordorf, D. (2010). Continued growth of the impact factors of MDPI open access journals. *Molecules, 15*(6), 4450–4451. https://doi.org/10.3390/molecules15064450.

Schwarz, G. J., & Kennicutt, R. C. (2004). Demographic and citation trends in astrophysical journal papers and preprints. *Bulletin of the American Astronomical Society*, *36*, 1654–1663. http://arxiv.org/abs/astro-ph/0411275.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628–638. https://doi.org/10.1002/(SICI)1097-4571(199210)43:9%3c628:AID-ASI5%3e3.0.CO;2-0.

Snijder, R. (2010). The profits of free books: An experiment to measure the impact of open access publishing. *Learned Publishing, 23*(4), 293–301. https://doi.org/10.1087/20100403.

Snijder, R. (2016). Revisiting an open access monograph experiment: Measuring citations and tweets 5 years later. *Scientometrics, 109*(3), 1855–1875. https://doi.org/10.1007/s11192-016-2160-6.

Solomon, D. J., Laakso, M., & Björk, B.-C. (2013). A longitudinal comparison of citation rates and growth among open access journals. *Journal of Informetrics, 7*(3), 642–650. https://doi.org/10.1016/j.joi.2013.03.008.

Sotudeh, H., Ghasempour, Z., & Yaghtin, M. (2015). The citation advantage of author-pays model: The case of Springer and Elsevier OA journals. *Scientometrics, 104*(2), 581–608. https://doi.org/10.1007/s11192-015-1607-5.

Sotudeh, H., & Horri, A. (2007). The citation performance of open access journals: A disciplinary investigation of citation distribution models. *Journal of the American Society for Information Science and Technology, 58*(13), 2145–2156. https://doi.org/10.1002/asi.20676.

Swan, A. (2010). *The Open Access citation advantage: Studies and results to date*. https://eprints.soton.ac.uk/id/eprint/268516. Accessed 10 April 2019

Teplitskiy, M., Lu, G., & Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology, 68*(9), 2116–2127. https://doi.org/10.1002/asi.23687.

Turk, N. (2008). Citation impact of open access journals. *New Library World, 109*(1–2), 65–74. https://doi.org/10.1108/03074800810846010.

Valderrama-Zurián, J.-C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics, 9*(3), 570–576. https://doi.org/10.1016/j.joi.2015.05.002.

Wagner, A. B. (2010). Open access citation advantage: An annotated bibliography. *Issues in Science and Technology Librarianship, Winter.*. https://doi.org/10.5062/F4Q81B0W.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics, 10*(2), 365–391. https://doi.org/10.1016/j.joi.2016.02.007.

Wang, X., Cui, Y., Xu, S., & Hu, Z. (2018). The state and evolution of Gold open access: A country and discipline level analysis. *Aslib Journal of Information Management, 70*(5), 573–584. https://doi.org/10.1108/AJIM-02-2018-0023.

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics, 103*(2), 555–564. https://doi.org/10.1007/s11192-015-1547-0.

Wray, K. B. (2016a). No new evidence for a citation benefit for author-pay open access publications in the social sciences and humanities. *Scientometrics, 106*(3), 1031–1035. https://doi.org/10.1007/s11192-016-1833-5.

Wray, K. B. (2016b). Still no new evidence: Author-pay open access in the social sciences and humanities. *Scientometrics, 107*(3), 1527–1529. https://doi.org/10.1007/s11192-016-1907-4.

Wren, J. D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ, 330*(7500), 1128. https://doi.org/10.1136/bmj.38422.611736.E0.

Xia, J., Myers, R. L., & Wilhoite, S. K. (2011). Multiple open access availability and citation impact. *Journal of Information Science, 37*(1), 19–28. https://doi.org/10.1177/0165551510389358.

Xia, J., & Nakanishi, K. (2012). Self-selection and the citation advantage of open access articles. *Online Information Review, 36*(1), 40–51. https://doi.org/10.1108/14684521211206953.

Zhang, Y. (2006). The effect of open access on citation impact: A comparison study based on web citation analysis. *Libri, 56*(3), 145–156. https://doi.org/10.1515/LIBR.2006.145.

Zhang, L., & Watson, E. M. (2017). Measuring the impact of gold and green open access. *The Journal of Academic Librarianship, 43*(4), 337–345. https://doi.org/10.1016/j.acalib.2017.06.004.