

Does open access to academic research help small, science-based companies?

SME access to
academic
research

95

ElHassan ElSabry

Mirqah for Scientific Research and Management Consulting, Cairo, Egypt, and

Koichi Sumikura

National Graduate Institute for Policy Studies, Tokyo, Japan

Received 13 April 2020
Revised 16 August 2020
Accepted 20 August 2020

Abstract

Purpose – This study investigates the extent to which a company's usage of open access (OA) literature for R&D activities depends on its size. The authors' assumption is that smaller pharmaceutical companies have less access to (usually expensive) journal subscriptions.

Design/methodology/approach – A fixed-effect Poisson model was used to study a panel dataset of USPTO pharmaceutical company patents. The dependent variable is the count of citations to OA resources in a given company patent.

Findings – Results support current anecdotal evidence that many SMEs suffer from high journal prices.

Originality/value – This result justifies the assumption made by policymakers about the potentially positive impact OA mandates have on national innovation activity. It was also shown that collaborating with universities can be a potential coping mechanism for companies that struggle to gain access to the journals they need. In addition to the novelty of its findings, this study introduces a new way to study the impact of OA in nonacademic contexts.

Keywords Knowledge transfer, Open innovation, Open access mandates, Science linkage

Paper type Research paper

Introduction

The importance of academic research for innovation

It is required for a company to cultivate an ability to absorb and adapt new and external information to increase the propensity to realize innovation for commercial application known specifically as “absorptive capacity” (Cohen and Levinthal, 1990). Conducting basic research internally can contribute to a company's efforts to improve its absorptive capacity, allowing it to broaden its own base of knowledge, discover and understand external sources of new knowledge and tie these activities to the development of new products and benefits. Thus, for companies, to what extent does external knowledge generated from basic academic research results contribute to company innovation?

A positive correlation between university research expenditure and corporate patent applications was shown by analyzing chronological data from different US states (Jaffe, 1989). It was also shown that published research in particular is among the key channels through which universities influence industrial R&D (Cohen *et al.*, 2002). According to corporate questionnaire surveys, without the results gleaned from academic research,

© ElHassan ElSabry and Koichi Sumikura. Published in *Journal of Industry-University Collaboration*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

The authors would like to thank Roberto Leon-Gonzales of GRIPS for his support and advice throughout this project.



13–15% of new products would not have been developed, or the appearance of these products would have been delayed considerably (Mansfield 1991, 1998).

Measuring “Science Linkage”

Apart from surveying companies about the importance of academic research to their R&D, researchers have attempted to develop more concrete methods to explore and assess this phenomena (Carpenter *et al.*, 1980). These efforts resulted in the creation of a number of indicators many of which is based on the analysis of patent and journal paper data. One indicator gained much attention in the literature. That is the concept of “Science Linkage” (SL), initially introduced by Francis Narin (1991). It is most commonly defined as “the average number of science papers referenced on the front page of the company’s patents” (Narin, 2000). This refers to the “References Cited” section in the typical United States Patent and Trademark Office (USPTO) patent document. It includes citations to US patents, foreign patents and “other publications”, which is usually referred as nonpatent literature (NPL). NPL includes anything from citations to journals articles to letters of communication during the patent examination process. The SL indicator was widely adopted in the literature to study different issues. For example, it was used to study the impact of publicly funded research on innovation, demonstrating that 73% of the academic papers cited in patents filed by US companies were produced by university or other public research institutions (Narin *et al.*, 1997), indicating that academic papers are an important information source for industry and thus access to them is crucial for R&D activities in industry.

It is believed that there is a higher probability that the knowledge from basic academic research is applied in the field of drug and medical products industry (Mansfield 1991, 1998). The proportion of products that would not have been developed if not for the results of academic research was estimated at an average of 15% across all industries between the years 1986 and 1994, whereas the figure was 31% for drugs and medical products, the highest percentage of all industries examined in the surveys (Mansfield 1991, 1998). Also according to survey results (Klevorick *et al.*, 1995), the pharmaceutical industry is where science is most entwined with business. The idea that pharmaceutical industry has the highest degree of SL (among all industries) is probably as old as the concept of SL itself (Narin *et al.*, 1992). However, recent SL studies still confirmed this pattern. For example, a study that analyzed NPL in patents granted by the USPTO over a 20-year period (1993–2012) showed that drug and medical related patents cited the highest amounts of science-based NPL (meaning journals, conference papers and books) (Sung *et al.*, 2015). Another study concluded that “drug patents rely heavily on knowledge derived from scientific research” based on the finding that 81.4% of patents in their sample cited at least one journal paper (Du *et al.*, 2019). It was also shown that pharmaceutical patents have seen the highest rate of increase of citations to academic papers since 1975 (Jefferson *et al.*, 2018).

According to these studies, it is clear that there is a strong connection between academic research and innovation, especially with respect to companies in the pharmaceutical industry. However, this raises the concern of how these companies (to whom academic research is important) access this academic knowledge. In this regard, firm size is a particularly interesting factor to study (Roper *et al.*, 2017) given the limitation on small firms to benefit from external sources of innovation.

Rising cost of knowledge and the open access movement

The huge boom in the number of researchers and universities worldwide after the Second World War, partly because of the unprecedented levels of government funding for science and partly due to the need of establishing nationally-oriented universities in newly independent states, has called for an accelerated expansion in the scholarly communication

enterprise (Price, 1963). This was mainly reflected in the takeover of the system by commercial publishers at the expense of traditional society publishers (Morrison, 2012). The introduction of financial interests to the system (as opposed to traditional motives like the pursuit of knowledge and intellectual merit) combined with the diffusion of neoliberal ideals in higher education in general has manifested in the continued rise of journal subscriptions prices, eventually leading to what became known as the “serials crisis of the 1990s”. At that time, the rate of price increase was so high (more than three times the consumer price index according to some estimates (Dingley, 2005)) that libraries around the world, especially in North America, started to cancel many of their subscriptions. In its report about the issue, the UK Office of Fair Trading noted that “there a number of features of this market that might militate against the operation of normal competitive market forces” (Office of Fair Trading, 2002). It is no wonder that the top 100 publishers own about two-thirds of the journals, while the top ten alone own 45% of them (Mark Ware Consulting 2009). Experts also estimate profit margins in the industry to be usually in the 20–30% range (Van Noorden, 2013).

In response to this trend (and taking advantage of the concurrent expansion of the world wide web), a group of researchers and librarians started to coordinate their efforts in what came to be called the “Open Access Movement”. The Budapest Open Access Declaration in 2002 can be seen as the first major milestone of the movement. It provided both a definition of for open access (OA) and a roadmap for how the movement should go about to achieve its goals. Since that time, lots of discussions, declarations, workshops, conferences and lobbying have taken place in support of (or in opposition to) open access.

For research to be OA, two kinds of barriers have to be eliminated (Suber, 2012). The first is the price barrier, which means that the research article should be accessible for anyone with Internet access at no extra cost. The second type is permission barriers, which are typically imposed by copyright (e.g. restrictions on printing, distribution and, more recently, text mining). Open access can be achieved in two main ways, either by publishing in OA journals (“gold OA”) or uploading copies of normally published papers in freely accessible archives (e.g. university repositories, PubMed Central and RePEc), which is “green OA”. One problematic aspect of the green route is that some of the journals enforce an embargo period (usually 6–12 months) on authors before they are allowed to share their papers in these open repositories. This can delay the diffusion of knowledge to its beneficiaries, which might create more harm to companies (under competitive pressures) than to academics (Carlino, 2001). There are a plenty of other “flavors” of open access (Willinsky, 2003), the most common of these other flavors is when a normally subscription journal offers its authors to make their individual articles OA in return for a fee (hybrid route). However, this hybrid model is very controversial. Many consider it “double-dipping” on the part of publishers because in many cases the journal subscription prices do decrease even when many individual articles are freely available (Prosser, 2015).

Literature review

Journal cost as a potential barrier to knowledge transfer

In attempting to express both linear and nonlinear characteristics of the knowledge transfer process, the interactive–recursive model of knowledge transfer suggests that this process happens as a temporal motion (linear part) in a three-dimensional space (nonlinear part) (Eckl, 2012). According to this model, the three dimensions of the knowledge transfer process are knowledge creation, knowledge diffusion and knowledge absorption. All three dimensions interact with each other, and the model specifies one determinant for the effectiveness of motion along each of the dimensions. For knowledge diffusion, the determinant of success is how effective the medium is in making the created knowledge known to the circle of its potential users. This is why it is important to identify the extent to which the published academic literature is successful in reaching all potential knowledge users.

Other attempts to identify barriers to knowledge transfer have considered financial constraints on companies as a possible barrier. Insufficient resources allocated to R&D were one of the seven barriers to knowledge transfer identified by [Irwin and More, \(1991\)](#). When faced with limited budgets, it is unlikely that R&D managers will give precedence to expensive journal subscriptions at the expense of other more vital resources needed for research (e.g. personnel or equipment).

Difficulties facing SMEs in accessing the academic literature

Multiple sources have voiced the concern about an ongoing crisis in access to the literature by small and medium enterprises (SMEs) ([Lyman, 2011](#); [CIBER, 2011](#); [Houghton et al., 2011](#); [Mark Ware Consulting 2009](#)). Apart from anecdotal evidence, five studies were conducted in attempt to better understand this issue. They were based on surveying companies in Denmark, Japan and the UK.

In two of the studies, the percentage of British SMEs claiming to have easy access to the literature was smaller than that of large companies. In one study ([Mark Ware Consulting, 2009](#)), only 71% of SMEs claimed so (compared to 86% of large companies and 94% of universities). This was controlled for companies' need for access. The other study ([CIBER, 2011](#)) confirmed the same trend and showed that 85% of researchers in industry and commerce have experienced a recent (unresolved) access problem, compared to only 44% of researchers in academia. Over one-third of researchers in industry (38% in SMEs and 35% in large companies) reported visiting a local public library in attempt to access needed research. On the other hand, OA research was the third most common way to access the literature (following personal and corporate subscriptions). The third study (also in the UK) used extensive interviews with representatives of different businesses in order to identify the potential OA research can have on fulfilling their research needs ([Parsons et al., 2011](#)).

Similar to the finding from the second study about pay-per-view access, the third study showed that OA is never systematically used as a way of accessing the needed literature. It is only encountered by coincidence while trying to locate the relevant literature. Participants identified other barriers for using the academic literature like the irrelevance of academic research to industry or the lack of skills needed to locate OA versions of the needed research. For some, the main benefit for OA research was the ability to easily scan large amounts of literature to identify potential collaborators from academia. In other words, OA publications are not used as a medium of knowledge transfer *per se*.

In Denmark, while 64% of survey respondents holding research roles in their companies stated that access to research is essential for the business, over half of all respondents have experienced difficulties in accessing the research they need ([Houghton et al., 2011](#)). In Japan, it was reported there was an increase between 2008 and 2011 in the percentage of companies that cutback on journal subscriptions due to budget limitations ([Abe et al., 2011](#)). This appears to be more common among small companies because, over the same period, the percentage of companies with 100 or more journal subscriptions has increased. Results from all of these surveys call for further investigation of what academic resource companies actually use and the extent to which analyzing usage data will confirm claims about difficulties in accessing the literature. The question this study aims to address is to what extent the size of a company contributes to these claimed access difficulties.

Hypothesis

Building on evidence about the negative impact of high journal prices on SMEs and how important the academic literature is for industrial research (especially in knowledge-intensive industries ([Parsons et al. 2011](#))), this study attempts to understand how company size relates to its dependence on freely accessible publications. The main hypothesis is that *small*

companies use more OA resources than large companies. To test this hypothesis, a dataset was constructed by matching citations to OA resources inside patents of US pharmaceutical companies to the sizes of companies which own them. Following the Frascati Manual (OECD, 2015), which offers authoritative guidelines to collect R&D statistics, the number of employees is used as a measure of company size. The assumption is that the company size is a measure for its ability to subscribe to journals. Therefore, small companies are expected to cite more OA resources in their patents than large companies because of their inability to subscribe to all the journals they need.

Research design

Data

To test this hypothesis, American pharmaceutical companies were selected owing to their large world share (around 50%) in drug patents (Friedman 2017), combined with the fact that the US is the largest market for pharmaceutical innovations (Du *et al.*, 2019). Firm-level data were extracted from Bureau van Dijk's Orbis database for all 1,109 US-based companies working in the pharmaceutical sector. This number covers companies designated in ORBIS database as "small", "medium-sized" and "large" but excludes those designated as "very large" (338 companies). There are three reasons why this exclusion is justified. First, this subset of companies collectively own more patents than all other 1,109 companies in the three other categories. Given the dependence of this study on manual labor to ensure accuracy in citation data matching (more details below), including these patents would have made the analysis process significantly more complicated, if not impossible. Second, assuming patents from these companies were included (whereby they would account for over half of patents in the dataset), it could potentially bias the results to reflect the journal citation behavior of very large companies. Third, and perhaps most important, the included companies had a size range which covers all but two of the nine company size categories proposed by the Frascati Manual (OECD, 2015). Only companies with zero employees or those with more than 5,000 are not represented in the used data. Therefore, included companies were judged to be enough for the purpose of the study.

Collectively, these 1,109 companies owned 8,968 patents granted by different patent authorities worldwide. In this study, only patents awarded by the USPTO were considered. These were 2,549 patents owned by 600 companies. All the numbers above are for patents granted between the years 2005 and 2014. Firm size data from ORBIS was available for a 10-year period (2004–2013). Matching patents to company size was done so that each patent is paired with its owner company size one year before the patent was granted. This would approximately reflect the company size around the time when patent examination was taking place. Selecting the company size at one year ahead of patent grant was judged to be more appropriate than the company size at the time of patent application. This is because applicants are allowed by the USPTO to add citations during the patent examination process.

Orbis is one of the most comprehensive business intelligence databases available. It has been common to use it recently in innovation studies (see for example (Nemethova *et al.*, 2019; Bertoni *et al.*, 2019; Colombo *et al.*, 2019; Knockaert *et al.* 2019)). Besides its extensive coverage of private companies, the data are well curated, especially with regards to harmonization of company names. The database version used in this study covered over 140 million companies worldwide. However, missing company size data was a general feature of the database, especially for private companies. Only around one-third of possible entries for countries in our study (one per year per company for ten years) were available in Orbis. Over half of the 2,549 patents could not be matched to the size of their respective owner company. In trying to remedy this problem, an attempt was made to interpolate some of the missing data. The majority of companies in our sample had five or more data points missing. For cases where the

number of employees was missing at one year but available for one year before and one year after, the arithmetic mean was calculated and taken to be the company size at middle year. Using this method, 153 records were added to the original 1,104 patent-company size pairs used in this study (a 14% increase). [Table 1](#) below shows how adding the interpolated data influence sample representativeness.

Disambiguation of NPL citations

From Orbis, data retrieved were as follows: company name, owner group (if any), company size (number of employees), total number of owned patents in addition to the classification and the identifier for each one of these patents. Only granted patents were counted. The USPTO database was then used to confirm the owner company and the grant year. Data on coowners of the patent (if any) and citations to the NPL was collected from the USPTO. NPL citations (starting in year 2001) are marked to identify whether the citation was made by the applicant or by the examiner during the examination process ([Cotropia et al., 2013](#)). Only applicant-provided citation was considered for the purpose of this study.

After extracting NPL citations for all patents in the dataset (44,087 citations), keyword search was used to exclude the majority of citations to nonjournal publications (e.g. correspondence, textbooks and reports). Around 10,000 nonjournal citations were identified and excluded. Following this, computer-assisted manual search was used to extract journal names from the remaining citations. NPL citations are provided in free text format, unparsed. Furthermore, the USPTO does not require applicants to cite journal papers in any specific citation style, meaning that applicants have the freedom to describe the cited paper in any format they choose. This is in addition to different ways to abbreviate journal names and nonstandard uses of punctuation marks. These factors made it difficult to use citation parsing software (assuming the results will not be reliable and will need further manual confirmation) and necessitated the tedious task of manual extraction of journal names ([Fedoryszak et al., 2013](#)). Finally, journal names could be identified within 33,216 citations to journal articles.

After identifying which NPL citations were journal citations, the next step was to identify which of these journals were OA at the time the patent was being examined. This was done by consulting all journal websites to which the 33,216 citations belonged. In some cases, it was also necessary to do an investigation of the journal's online history (using the internet Archive's Wayback Machine). Current OA journal databases (e.g. DOAJ) could not be used to determine if the journal is OA because what matters is whether the journal was OA at the time the paper was cited.

On completing the process of NPL citation disambiguation, it was found that 239 patents out of the 1,257 patents matched with company size data (19%) did not cite any academic journals. They consist of 179 patents from the original dataset in addition to 60 patents from the interpolated data and where excluded from further analysis. The reason for exclusion is to ensure that cases with zero OA citations can be (at least partly) attributed to the company's financial ability to access subscription journals not to lack of interest in academic research in

Table 1.

Data on how adding the interpolated data influences sample representativeness

Number of employees	Before matching (<i>n</i> = 2,549)	Matched data (<i>n</i> = 825)	Matched data after interpolation (<i>n</i> = 918)	Share of all journal citations	Share of OA citations
1–9	18.8%	16.6%	17.8%	18.0%	21.7%
10–49	23.5%	22.2%	23.0%	27.4%	20.5%
50–249	35.9%	43.3%	41.8%	37.1%	46.1%
250+	21.9%	17.8%	17.3%	17.5%	11.8%

general. This means that the data eventually used for the regression consisted of 918 patents, 825 from the original data complimented by 93 patents after interpolation.

Dependent variable

The dependent variable is the count of citations to OA resources mentioned in the NPL section of the patent document. Another measure was considered at the early stages of this study. That is the ratio of OA citations to all journal citations in the patent. But it was later dropped in favor of using a count variable that was suitable to model using Poisson regression, given the highly skewed nature of the data (with most patents citing very few or no OA publications). To avoid the bias from using counts of OA citations not their ratio, the total number of citations to journals in the patent was introduced as a control variable in the model.

A cited reference was considered OA if it belongs to a journal that was established as OA (“born open”) or if it was published in a subscription journal that later became OA. This was the case whether the paper was published before (“before open”) or after the conversion (“after open”), given that the majority of journals that convert to OA make their archives openly available. The important factor was whether the conversion took place prior to the patent examination period (defined as one year before patent grant in this study).

Citations where the applicant specifically mentioned that they cite the paper abstract or those where an abstract service is cited (e.g. chemical abstracts) were considered to be citations to OA papers, given that all online journals make their abstracts accessible at no cost. These (citations to abstracts) accounted for the majority of OA citations (1,330 citations, 61% of OA citations and 4.6% of total journal citations).

Before interpolation, there was a total of 2,177 OA citations making up about 8% of all journal citations (Figure 1). The ratio was almost the same after adding OA citations from patents analyzed after interpolation.

Explanatory variables

- (1) *Company size* is the main variable of interest in this analysis, measured by the number of employees.
- (2) *Patent class* is a dummy variable which takes the value one if the patent is classified under A61, C07, C08 and C12 (CPC system) and zero otherwise. Although all companies in the dataset are pharmaceutical companies, not all patents are classified as such. Many pharmaceutical companies own patents that belong to other fields of research (e.g. polymers). In general, if the company allocates resources to journal subscriptions, the assumption is that it will give preference to biomedical journals. This means that company researchers are more likely to seek OA journals when their invention is outside the main field of the company.

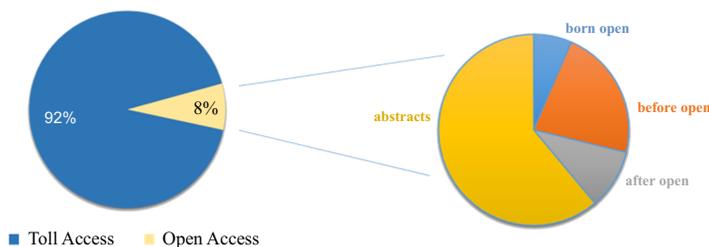


Figure 1. Breakdown of OA citations in the NPL section of pharmaceutical patents

- (3) *University collaborator* is a dummy variable that equals one if the company coowns the patent with a university (or some research institution/hospital) and zero otherwise. If the patent at the hand was a result of research collaboration, it is possible that sharing journal access privileges with collaborators would decrease the need to seek OA alternatives.
- (4) *Company collaborator* is similar to university collaborator (following the same argument of resource sharing) but takes the value of one in case of patent coownership with another company.

In addition, the overall technological capability of the company is a good determinant of its research activity. Research active companies would presumably subscribe to more journals than other companies, which means they would rely less on OA. A measure of such technological capability (counting all patents owned by the company over the ten-year period) was initially used in this study but later dropped in favor of company fixed effects.

Control variables

Two variables assumed to influence the count of OA references in patent citations of pharmaceutical companies were included in the regression as control variables.

- (1) *Total journal citations* had to be accounted for because some patents tend to cite academic journals less than others depending on the novelty of the technology as well as other factors.
- (2) *DOAJ journals* (listed as of the year before the patent grant) is supposed to be a good proxy to the amount of openly available literature. DOAJ (Directory of OA Journals) is the industry standard for indexing OA journals and has been growing ever since the open access movement emerged (DOAJ). By adding this to the regression, the assumption is that the more OA articles become available, the more companies are likely to cite them

Descriptive statistics

Table 2 below provides some basic statistics for all variables in the data set. All are calculated for the full dataset (including interpolated data) with 918 records.

Descriptive analysis of the data shows an interesting trend regarding patent citations to OA resources. These citations have increased over the period of 2005–2013 by an average annual growth rate (AAGR) of 41%, rising twice as fast as patent citations to journals in general (both OA and not) with an AAGR of just 20%. This increase can also be observed at the level of patents. In other words, it is not just the number of OA references in a given patent that is increasing but the number of patents that cite OA references is itself increasing. This also happens at a faster rate (22% AAGR) than the rate of increase of patents citing journals in general (16%). The number of patents citing OA journals has sustained an average annual increase of 22% versus only 15% increase in patents citing journals in general.

In preparation for regression, the correlation matrix of all variables (Table 3) was checked. Four out of six coefficients of correlation (between the dependent variable and each one of the six independent variables) provide initial evidence in support of assumptions made in this study. These are coefficients of correlation of OA citations with employee count, total journal citations, university collaboration and DOAJ.

Regression model

The semicontinuous nature of the dataset (i.e. having a large number of zero values with the distribution of the remaining data heavily left skewed) calls for using a two-step (hurdle)

count model (Olsen and Schafer, 2001). The first step is a logit regression that determines the probability of citing at least one OA resource. The second step of the model is a zero-truncated Poisson regression that takes into consideration company fixed effects. The main purpose is to determine the relationship between the owner company size (one year before the patent is granted) and the number of OA resources cited in one of its patents. Using a Poisson model was a better fit to the skewness found in the data, with most patents not citing OA journals at all. The result of Hausmann test has shown preference (at the 0.05 significance level) for using fixed effects in the model as opposed to random effects, with company code as the panel variable.

Results

Company size as a determinant of OA citations

Consistent with the main hypothesis of this study, results show that the smaller the size of the company, the more dependent it is on OA research for its internal R&D leading to patents. This is evident in the negative sign of the coefficient for “employee count” in all models reported in Table 4, except Model-VI. A potential explanation for this effect is the crisis of soaring journal prices, whereby smaller companies struggle to buy subscription to all the journals they need (Lyman, 2011). They resort to citing OA journals, or even just abstracts of subscription journal papers.

Regarding the positive coefficient in Model-VI for “employee count”, it is difficult to make assertions regarding this case because adding company fixed effects essentially controls for company size as well for the dataset as a whole. Since “employee count” is a company-specific variable (albeit sometimes varies for patents granted in different years), its coefficient can be interpreted as effect of variation from the mean company size on the variation in counts of OA citations, which leaves room for complication given the small number of OA citations in general in the dataset.

Variable	Mean	SD	Min	Max
OA citations	2.583	6.936	0	92
Employee count	152.599	209.412	1	1,414
Total journal citations	36.183	51.656	1	353
Patent class	0.895	0.306	0	1
University collaborator	0.050	0.218	0	1
Company collaborator	0.036	0.186	0	1
DOAJ journals	1888.925	1,033.444	385	4,081

Table 2. Basic statistics for all variables in the data set

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) OA citations	1						
(2) Employee count	-0.0523	1					
(3) Total journal citations	0.4197	-0.0266	1				
(4) Patent class	0.0077	-0.0118	0.0902	1			
(5) University collaborator	-0.0179	-0.0644	0.0893	0.0622	1		
(6) Company collaborator	0.0023	0.0958	-0.0152	0.0086	-0.0444	1	
(7) DOAJ journals	0.1052	0.0103	0.0044	-0.1424	-0.0242	-0.0372	1

Table 3. The correlation matrix of all variables

Table 4.
The negative coefficient of variables using Poisson logit and zero-truncated Poisson models

	Poisson model I	Logit model II	III	Zero-truncated Poisson model IV	V	VI
Employee count	-0.0004* (0.0002)	-0.0007 (0.0038)	-0.001*** (0.0004)		-0.0006*** (0.0002)	0.0023*** (0.0011)
Total journal citations	0.0095*** (0.0009)	0.0615*** (0.0085)		0.0076*** (0.0009)	0.0074*** (0.0009)	0.0117*** (0.0025)
Patent class	-0.0951 (0.3671)	0.9160** (0.4627)		-0.5603* (0.3335)	-0.5524* (0.3337)	-0.1384 (0.0977)
University collaborator	-0.6369*** (0.2332)	-0.1418 (0.7662)		-0.5949*** (0.2293)	-0.6024** (0.2364)	-0.4753*** (0.119232)
Company collaborator	0.0733 (0.2193)	0.1692 (0.5857)		0.0715 (0.2070)	0.1883 (0.2057)	0.2119 (0.2196)
DOAJ journals	0.0003*** (0.0001)	0.0003 (0.0002)		0.0003*** (0.000)	0.0003*** (0.000)	0.0002*** (0.000)
Constant	-0.0037 (0.3630)	n/a	1.6508*** (0.1094)	0.9750*** (0.3536)	1.0785*** (0.3558)	n/a
Company FE	N	Y	N	N	N	Y
Prob > χ^2	0.000	0.000	0.0035	0.000	0.000	0.000
MacFadden adj R^2	0.2406		0.0154	0.2218	0.2281	
Observations	918	918	517	517	517	517

Note(s): (robust) standard errors are in parentheses, *** $p < 0.001$, ** $p < 0.01$ and * $p < 0.05$

Impact of R&D collaborations

A very important finding from [Table 4](#) is that regarding companies collaboration with universities that result in coowned patents. Analysis shows that in almost all cases, there is a highly significant negative coefficient for the variable “university collaborator”. This implies that university–industry collaboration can be beneficial for companies in the sense that (because of the large journal databases universities have access to) they will be less dependent on the OA literature to explore prior research. In this regard, collaboration with universities can be viewed as potential coping mechanisms for companies (especially small ones) that struggle to access the literature they need.

Contrary to what was initially assumed, collaboration with other companies seems to have no significant effect on citing OA journals. This can potentially be explained by the fact that pharmaceutical companies (especially those engaging in collaboration) might have very similar research interests. This can result in small variation in the pool journals the two companies have joint access leading to higher dependence on OA resources. On the other hand, in cases where a company collaborates to complement its research capabilities, it is more likely that the other company will have different interests and (consequently) subscribe to different journals, which will expand the pool of journals available to both companies and calls for the less usage of OA literature. These two opposing trends might have resulted in the inconclusive relationship between having a “company collaborator” and usage of OA resources.

Patent class relevance

It has been established that most of pharmaceutical company patents fall within a very narrow subset of patent classes ([Narin et al., 1987](#)). While there is a distinction between companies that are “drug-dependent” and companies have a portfolio of different products, it is fair to assume that the interest of pharmaceutical companies in general falls within the life science/biochemical realm. As mentioned before, the definition of a “Relevant Patent Class” was restricted to only four CPC patent classification subsections. These cover medical and veterinary science (A61), organic chemistry (C07), organic macromolecular compounds (C08) and biochemistry, microbiology, enzymology and genetic engineering (C12). The choice of these general areas is supposed to strike a balance between the narrow span of pharmaceutical patent classes and the wide scope that some biomedical journals tend to have.

The negative value of the “patent class” coefficient reported in [Table 4](#) suggests that patents outside the scope in which pharmaceutical companies operate tend to have a higher count of citations to OA resources. This is in agreement with the initial assumption that whenever resources are available to buy journal subscriptions, a company will use them to buy journals that publish research in the company’s narrow scope of interest. This consequently leads to higher dependence (or at least higher propensity to use) OA journals.

The fact that this same coefficient is not significant after adding company fixed effects to the model can be explained by the influence large companies (that usually own a large number of patents) have on the fixed effect regression outcome, given that these companies have a higher probability of owning nonpharmaceutical patents.

Discussion

There are three important considerations in interpreting these results that are also necessary to estimate the potential for their generalizability.

First, using manual methods in the disambiguation of NPL citations (as opposed to parsing algorithms) proved to be extremely more rewarding in terms of avoiding data loss. Among the 44,087 NPL citations handled in this study, only 79 citations (less than 0.2%)

could not be attributed to a particular source. This is in stark contrast to the other studies that use computer algorithms to perform the matching process. Needless to say, manual matching has the disadvantage of being labor intensive and time consuming. However, this study will supposedly reflect a more accurate estimation of Science Linkage effects.

Second, as mentioned before, only abstracts and articles published in OA journals were counted as OA for the purpose of this analysis, which would naturally result in an underestimation of the counts of OA citations because of neglecting other variations of OA papers. For example, OA papers published in hybrid journals (those combining OA and subscription models) will not be counted as OA in this study because OA status is determined at the journal level. Hybrid journals in this case do not count as OA. Another, perhaps more significant portion of papers are those published as “green OA”. These are (as explained above) papers published in subscription journals, but some version of the paper is deposited in a preprint archive or some other online database. This is especially important in the context of this study because pharmaceutical and biotechnology companies can benefit from the existence of the US federally-funded repository PubMed Central, which makes available huge amounts of green OA papers. Such papers cannot be counted as OA if matching is done at the journal level, like we did in this study. In this regard, a recent study (Bryan and Ozcan, 2018) has attempted to investigate the impact of green OA in particular on access to knowledge by firms. Indeed, this is a very interesting topic to pursue in future research.

Third, the issue of how to determine the quality of journals is indeed a very debatable one (Brembs *et al.*, 2013). However, it can still be argued that the relative short age of all OA journals might put them at a disadvantage with traditional journals that have established good reputation over their long history. Therefore, they might not have gained enough popularity (judging by the minority of citations to “born open” journals) among industry researchers to be extensively consulted in their research activities.

Conclusion

Results of this study show that OA research might be an efficient way used by small firms to overcome the barrier to knowledge transfer created by high journal subscription costs. This is one type of barrier to knowledge transfer that has not been studied in the prior literature. Failing to reject the initial hypothesis (that smaller companies utilize OA articles more than large companies) supports current anecdotal evidence that many SMEs suffer from high journal prices (Lyman, 2011). It also supports the assumption made by many policymakers about the potentially positive impact OA mandates have on national innovation activity (ElSabry 2017a).

This is an important result because more and more governments and research funding agencies are obliged to take sides on the OA debate. On the one hand, they are under pressure to “mandate” OA publishing on their grantees (and pay for it) to ensure a more equitable research environment and to make it easier to create new knowledge by building on prior efforts in the literature. On the other hand, critics argue that this is a breach on academic freedom as researchers should have the only say in where to publish their research. A careful look at these two opposing arguments show that both fall within the realm of impact of OA on the research community. What this study and a number of others recently (ElSabry, 2017b) conclude is that the potential impact of OA transcends the boundaries of “the academic community”. More groups (SMEs in the pharmaceutical industry, for example) have a stake in whether OA publishing becomes main mode of research communication.

In addition to these findings, this study has introduced a new way to study the impact of OA research in nonacademic contexts. Examining citations to OA research is not restricted to patent documents. The same can be done for other types of documents like clinical guidelines or government reports. The effect of resources available for the citing body (turnover for

companies, budget for government research units, etc.) can be a good indicator for their propensity to cite OA journals versus citing journals that require subscription.

SME access to
academic
research

References

- Abe, K., Yaguchi, M. and Shimase, S. (2011), “企業等における電子ジャーナル等の利用動向調査”, *Journal of Information Processing and Management*, Vol. 54 No. 7, pp. 400-405, doi: [10.1241/johokanri.54.400](https://doi.org/10.1241/johokanri.54.400).
- Bertoni, F., Martí, J. and Reverte, C. (2019), “The impact of government-supported participative loans on the growth of entrepreneurial ventures”, *Research Policy*, Vol. 48 No. 1, pp. 371-384, doi: [10.1016/j.respol.2018.09.006](https://doi.org/10.1016/j.respol.2018.09.006).
- Brembs, B., Button, K. and Munafo, M. (2013), “Deep impact: unintended consequences of journal rank”, *Frontiers in Human Neuroscience*, Vol. 7, doi: [10.3389/fnhum.2013.00291](https://doi.org/10.3389/fnhum.2013.00291).
- Bryan, K.A. and Ozcan, Y. (2018), *The Impact of Open Access Mandates on Invention*, Mimeo, Toronto, available at: <http://www.kevinbryanecon.com/ManagementOADraftFinal.pdf>.
- Carlino, G.A. (2001), “Knowledge spillovers: cities’ role in the new economy”, Business Review Q4, pp. 17-26, available at: <https://ideas.repec.org/a/fip/fedpbr/y2001iq4p17-26.html>.
- Carpenter, M.P., Cooper, M. and Narin, F. (1980), “Linkage between basic research literature and patents”, *Research Management*, Vol. 23 No. 2, pp. 30-35, doi: [10.1080/00345334.1980.11756595](https://doi.org/10.1080/00345334.1980.11756595).
- Cotropia, C.A., Lemley, M.A. and Sampat, B. (2013), “Do applicant patent citations matter?”, *Research Policy*, Vol. 42 No. 4, pp. 844-854, doi: [10.1016/j.respol.2013.01.003](https://doi.org/10.1016/j.respol.2013.01.003).
- CIBER (2011), “Access to scholarly content: gaps and barriers”, *Research Information Network (RIN), Publishing Research Consortium (prc) and JISC*, available at: http://www.rin.ac.uk/system/files/attachments/gaps_final_report_low_res.pdf.
- Cohen, W.M. and Levinthal, D.A. (1990), “Absorptive capacity: a new perspective on learning and innovation”, *Administrative Science Quarterly*, Vol. 35 No. 1, pp. 128-152, doi: [10.2307/2393553](https://doi.org/10.2307/2393553).
- Cohen, W.M., Nelson, R.R. and Walsh, J.P. (2002), “Links and impacts: the influence of public research on industrial R&D”, *Management Science*, Vol. 48 No. 1, pp. 1-23, doi: [10.1287/mnsc.48.1.1.14273](https://doi.org/10.1287/mnsc.48.1.1.14273).
- Colombo, M.G., D’Adda, D. and Quas, A. (2019), “The geography of venture capital and entrepreneurial ventures’ demand for external equity”, *Research Policy*, Vol. 48 No. 5, pp. 1150-1170, doi: [10.1016/j.respol.2018.12.004](https://doi.org/10.1016/j.respol.2018.12.004).
- Dingley, B.U.S. (2005), “Periodical prices-2005”, *USA Periodical Price Index*, pp. 1-16, 2005.
- Du, J., Li, P., Guo, Q. and Tang, X. (2019), “Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis”, *Journal of Informetrics*, Vol. 13 No. 1, pp. 132-148, doi: [10.1016/j.joi.2018.12.004](https://doi.org/10.1016/j.joi.2018.12.004).
- Eckl, V. (2012), “Barriers of knowledge transfer”, Presented at the DRUID 2012, CBS, Copenhagen, available at: <https://pdfs.semanticscholar.org/efe4/c60dc0ee0fedd9b582aeea6e1c3525e7e012.pdf>.
- ElSabry, E. (2017a), “Claims about benefits of open access to society (beyond academia)”, *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, IOS Press, pp. 34-43, available at: <http://ebooks.iospress.nl/publication/46640>.
- ElSabry, E. (2017b), “Who needs access to research? Exploring the societal impact of open access”, *Revue Française Des Sciences de l’information et de La Communication*, No. 11, doi: [10.4000/rfsic.3271](https://doi.org/10.4000/rfsic.3271).
- Fedoryszak, M., Bolikowski, Ł., Tkaczyk, D. and Wojciechowski, K. (2013), “Methodology for evaluating citation parsing and matching”, in Bembenik, R., Skonieczny, L., Rybinski, H., Kryszkiewicz, M. and Niezgodka, M. (Eds), *Intelligent Tools for Building a Scientific Information Platform: Advanced Architectures and Solutions*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 145-154, doi: [10.1007/978-3-642-35647-6_11](https://doi.org/10.1007/978-3-642-35647-6_11).
- Houghton, J., Swan, A. and Brown, S. (2011), *Access to Research and Technical Information in Denmark (Report)*, Danish Agency for STI and Denmark’s Electronic Research Library, available at: <https://eprints.soton.ac.uk/272603/>.

- Irwin, H. and More, E. (1991), "Technology transfer and communication: lessons from silicon valley, route 128, carolina's research triangle and hi-tech Texas", *Journal of Information Science*, Vol. 17 No. 5, pp. 273-280, doi: [10.1177/016555159101700503](https://doi.org/10.1177/016555159101700503).
- Jaffe, A. (1989), "Real effects of academic research", *American Economic Review*, Vol. 79 No. 5, pp. 957-970, available at: https://econpapers.repec.org/article/aeaarev/v_3a79_3ay_3a1989_3ai_3a5_3ap_3a957-70.htm.
- Jefferson, O.A., Jaffe, A., Ashton, D., Warren, B., Koellhofer, D., Dulleck, U. and Jefferson, R.A. (2018), "Mapping the global influence of published research on industry and innovation", *Nature Biotechnology*, Vol. 36, pp. 31-39, doi: [10.1038/nbt.4049](https://doi.org/10.1038/nbt.4049).
- Klevatorick, A.K., Levin, R.C., Nelson, R.R. and Winter, S.G. (1995), "On the sources and significance of interindustry differences in technological opportunities", *Research Policy*, Vol. 24 No. 2, pp. 185-205, doi: [10.1016/0048-7333\(93\)00762-1](https://doi.org/10.1016/0048-7333(93)00762-1).
- Knockaert, M., Deschryvere, M. and Lecluyse, L. (2019), "The relationship between organizational interdependence and additionality obtained from innovation ecosystem participation", *Science and Public Policy*, doi: [10.1093/scipol/scz002](https://doi.org/10.1093/scipol/scz002).
- Lyman, S. (2011), "Industry access to the literature [Comments and Opinion]", *Nature Biotechnology*, Vol. 29, pp. 571-572, doi: [10.1038/nbt.1909](https://doi.org/10.1038/nbt.1909).
- Mansfield, E. (1991), "Academic research and industrial innovation", *Research Policy*, Vol. 20 No. 1, pp. 1-12, doi: [10.1016/0048-7333\(91\)90080-A](https://doi.org/10.1016/0048-7333(91)90080-A).
- Mansfield, E. (1998), "Academic research and industrial innovation: an update of empirical findings", *Research Policy*, Vol. 26 Nos 7-8, pp. 773-776, doi: [10.1016/S0048-7333\(97\)00043-7](https://doi.org/10.1016/S0048-7333(97)00043-7).
- Mark Ware Consulting Ltd (2009), *Access by UK Small and Medium-Sized Enterprises to Professional and Academic Information*, Publishing Research Consortium, available at: <http://publishingresearchconsortium.com/index.php/prc-documents/prc-research-projects/33-sme-access-research-report/file>.
- Morrison, H.G. (2012), *Freedom for Scholarship in the Internet Age*, (Thesis), Communication, Art & Technology: School of Communication. available at: <http://summit.sfu.ca/item/12537>.
- Narin, F. (1991), "Globalization of research, scholarly information, and patents—ten year trends", *The Serials Librarian*, Vol. 21 Nos 2-3, pp. 33-44, doi: [10.1300/J123v21n02_05](https://doi.org/10.1300/J123v21n02_05).
- Narin, F. (2020), *Tech-line Background Paper*, CHI Research, Haddon Heights, NJ, available at: https://www.academia.edu/747343/Tech-line_background_paper.
- Narin, F., Noma, E. and Perry, R. (1987), "Patents as indicators of corporate technological strength", *Research Policy*, Vol. 16 No. 2, pp. 143-155, doi: [10.1016/0048-7333\(87\)90028-X](https://doi.org/10.1016/0048-7333(87)90028-X).
- Narin, F., Albert, M.B. and Smith, V.M. (1992), "Technology indicators in strategic planning", *Science and Public Policy*, Vol. 19 No. 6, pp. 369-381, doi: [10.1093/spp/19.6.369](https://doi.org/10.1093/spp/19.6.369).
- Narin, F., Hamilton, K.S. and Olivastro, D. (1997), "The increasing linkage between U.S. technology and public science", *Research Policy*, Vol. 26 No. 3, pp. 317-330, doi: [10.1016/S0048-7333\(97\)00013-9](https://doi.org/10.1016/S0048-7333(97)00013-9).
- Nemethova, V., Siranova, M. and Sipikal, M. (2019), "Public support for firms in lagging regions—evaluation of innovation subsidy in Slovakia", *Science and Public Policy*. doi: [10.1093/scipol/scy046](https://doi.org/10.1093/scipol/scy046).
- OECD (2015), *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, the Measurement of Scientific, Technological and Innovation Activities*, OECD Publishing, available at: https://www.oecd-ilibrary.org/science-and-technology/frascati-manual-2015_9789264239012-en.
- Office of Fair Trading (2002), *The Market for Scientific, Technical and Medical Journals*, Office of Fair Trading, Great Britain.
- Olsen, M.K. and Schafer, J.L. (2001), "A two-Part Random-effects model for semicontinuous longitudinal data", *Journal of the American Statistical Association*, Vol. 96 No. 454, pp. 730-745, doi: [10.1198/016214501753168389](https://doi.org/10.1198/016214501753168389).

-
- Price, D.J. de S. (1963), *Little Science, Big Science*, Columbia University Press, New York.
- Prosser, D. (2015), *The Costs of Double Dipping*, available at: <https://www.rluk.ac.uk/the-costs-of-double-dipping/> (accessed 2 March 2019).
- Parsons, D., Willis, D. and Holland, J. (2011), "Benefits to the private sector of open access to higher education and scholarly research", *JISC*, p. 52.
- Roper, S., Love, J.H. and Bonner, K. (2017), "Firms' knowledge search and local knowledge externalities in innovation performance", *Research Policy*, Vol. 46 No. 1, pp. 43-56, doi: [10.1016/j.respol.2016.10.004](https://doi.org/10.1016/j.respol.2016.10.004).
- Suber, P. (2012), *Open Access*, MIT Press, Cambridge, MA.
- Sung, H.Y., Wang, C.C., Huang, M.H. and Chen, D.Z. (2015), "Measuring science-based science linkage and non-science-based linkage of patents through non-patent references", *Journal of Informetrics*, Vol. 9 No. 3, pp. 488-498, doi: [10.1016/j.joi.2015.04.004](https://doi.org/10.1016/j.joi.2015.04.004).
- Van Noorden, R. (2013), "Open access: the true cost of science publishing", *Nature News*, Vol. 495 No. 7442, p. 426, doi: [10.1038/495426a](https://doi.org/10.1038/495426a).
- Willinsky, J. (2003), "The nine flavours of open access scholarly publishing", *Journal of Postgraduate Medicine*, Vol. 49 No. 3, p. 263, available at: <http://www.jpjgmonline.com/article.asp?issn=0022-3859;year=2003;volume=49;issue=3;page=263;epage=267;aulast=Willinsky;type=0>.

Further reading

- Narin, F. and Olivastro, D. (1992), "Status report: linkage between technology and science", *Research Policy*, Vol. 21 No. 3, pp. 237-249, doi: [10.1016/0048-7333\(92\)90018-Y](https://doi.org/10.1016/0048-7333(92)90018-Y).

Corresponding author

ElHassan ElSabry can be contacted at: doc14153@grips.ac.jp