The funding factor: A cross-disciplinary examination of the association between research funding and citation impact

Erjia Yan[1*], Chaojiang Wu[2], Min Song[3]

[1]College of Computing and Informatics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, U.S.A.

[2]LeBow College of Business, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, U.S.A.

[3]Department of Library and Information Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu Seoul 03722, KOREA

*Corresponding author

Email: ey86@drexel.edu

Abstract

This paper intends to illuminate the relationship between science funding and citation impact in seven STEMM disciplines (science, technology, engineering, mathematics, and medicine). Using a regression model with Heckman bias correction, we find that funding has a positive, significant association with a paper's citations in STEMM fields. Further analyses show that this association is magnified by the factors of multiple authorship and multiple institutions. For funded papers in STEM, multi-author and multi-institution papers tend to receive even more citations than single-authored and single-institution papers; however, funded papers in Medicine received less gain in citation impact when either factor is considered. Based on the finding that funding support has a stronger association with citation impact when it is treated as a binary variable than as a count variable, this paper recommends the allocation of funding to researchers without active funding support, instead of giving awards to those with multiple funding supports at hand.

Introduction

Investments in science and technology have a significant impact on economic growth and society's well-being (Lane, 2009). Guided by this belief, governments worldwide have come to see research and development (R&D) expenditures as an imperative investment. The National Science Board has estimated that $1.67 trillion (purchasing power parity) was spent on R&D in 2013 globally (National Science Board, 2016). The vast scale of these investments necessitates an accountability assessment, an empirical test of the claim that such investments benefit knowledge production and innovation (Holton, 1978).

Kate Knuth, an environmental scientist and a former Minnesota House of Representative, argued in a recent article in *The Atlantic* "*[s]cientists bring a unique perspective in how they look at data and think about problems. They're trained to value evidence, and to change their minds in the face of evidence. Right now, in a lot of our governance, we have people who just say this is the way it is, in the face of huge evidence to the contrary. That makes it hard to make good policy* (Yong, 2017)." Good policies are always grounded in solid evidence and data. There are several U.S. government funding and data statistics, such as the National Center for Science and Engineering Statistics (NCSES) and Science and Technology for America's Reinvestment—Measuring the EffecTs of Research on Innovation, Competitiveness, and Science (STAR METRICS). In particular, STAR METRICS revolutionized science of science policy (SoSP) research by the use of evidence-based metrics. This is a significant improvement over the previously practiced "anecdotes or localized studies that are frequently written to justify" (Largent & Lane, 2012, p. 431). Several empirical studies were made available using this platform (Cragin, Nichols, Simon, & Watts, 2012; Lane & Bertuzzi, 2011; Lane, Owen-Smith, Rosen, & Weinberg, 2015; Sarli & Carpenter, 2014; Van Noorden, 2015; Weinberg et al., 2014). For instance, studies have looked into the impact of science funding on employment (Lane & Bertuzzi, 2010, 2011) and network visualizations have been used to illustrate the clustering of grants and faculty (Lane et al., 2015). Studies have also shown the differences in innovativeness of research communities (Boyack & Klavans, 2015), in workforce composition of different funding divisions of NSF and NIH, and in the geographic distribution of vendor and subaward expenditures (Weinberg et al., 2014).

These studies have revealed the long term impact of science investments as new knowledge gained, jobs created, and new economic activity encouraged (Cragin et al., 2012; Lane, 2009; Lane & Bertuzzi, 2011; Lane et al., 2015; Sarli & Carpenter, 2014; Van Noorden, 2015; Weinberg et al., 2014). The present study focuses on the short-term impact of funding, as measured by its association with research outputs. Prior efforts in this area have largely focused on data specific to an individual funding organization, or to a particular research domain. Studies of the former type have examined the funding impact of the National Institute on Aging (Boyack & Börner, 2003), Engineering and Physical Sciences Research Council (Ma, Mondragón, & Latora, 2015), Natural Sciences and Engineering Research Council of Canada (Fortin & Currie, 2013), National Natural Science Foundation of China (Wang, Liu, Ding, & Wang, 2011), Transdisciplinary Tobacco Use Research Center of the National Cancer Institute (Trochim, Marcus, Mâsse, Moser, & Weld, 2008), and National Cancer Institute of Canada (Campbell et al., 2010). A domain-specific approach has been applied to library and information science (Cronin & Shaw, 1999; Zhao, 2010), nanotechnology (Wang & Shapira, 2011), and biological chemistry (Rigby, 2013), among other fields.

Prior research has employed descriptive statistics and regression analysis to investigate the correlation between funding and research outputs, both within a domain (Wang & Shapira, 2011) and across funding agencies (Leydesdorff & Wagner, 2009; Wang et al., 2011). Scholars have found a significant but weak relationship between citation impact and funding supports (Fortin & Currie, 2013; Jacob & Lefgren, 2011; Rigby, 2013). These studies, however, have tended to be restricted in disciplinary scope; what is lacking is a cross-disciplinary examination of the data. The weak relationship between research outputs and funding supports within disciplines may disguise important cross-disciplinary trends in that relationship.

Moreover, the availability of funding acknowledgement data in the last few years makes it possible to investigate funding impacts in many fields in a previously unattainable degree of detail (Gök, Rigby, & Shapira, 2016). To that end, this study makes full use of the funding acknowledgement data provided by Clarivate's Web of Science database. This database allows us to include publications from different science domains, thus significantly extending the disciplinary breadth of this study's undertaking. The primary objective of this study is therefore to identify cross-disciplinary patterns in the relationship between funding and research outputs. Our results provide empirical evidence that will help to address the accountability issues inherent in science investment; such evidence may also catalyze the development of new practices in research funding assessment. On a yet broader level, this study advances our understanding of the role of funding in knowledge- and innovation-making and promotes a scientific basis for all phases of the research process.

Materials and Methods

Data

This study includes bibliographic data from seven disciplines, selected to ensure data representativeness across different domains of science, technology, engineering, and mathematics (STEM), and medicine. For ease of interpretation, we combined the STEM fields with medicine and formed STEMM, a variant of STEM; this broader category includes Astrophysics, Computer Science, Engineering, Environmental Studies, Mathematics, Medicine, and Nanotechnology[1].

To mitigate journal sampling bias, we selected five closely cited journals to represent each discipline. For each discipline, we first located a flagship journal. Flagship journals are typically easy to distinguish, for instance, the *New England Journal of Medicine* in medicine. Even though, sometimes, there may be more than one flagship journal in a discipline, based on the way of identifying top journals introduced below,

[1] In our original data collection, we also included several social science and humanity domains; however, during analyses, we realized that the percentage of papers with funding acknowledgement in these domains was rather inconsistent. We believe this is an artifact caused by the inconsistent coverage of the database, as pointed out by (Álvarez-Bornstein, Morillo, & Bordons, 2017; Tang, Hu, & Liu, 2017) and decided to only focus on STEMM domains.

all flagship journals can be readily captured. A flagship journal is used as a seed to find top five journals that cited by the seed the most using journal citation data from the Web of Science database. For each of the five journals, their top five most cited journals were also identified, resulting in 25 cited journal instances. Next, aggregating these instances, we obtained a list of the most cited journals by the five journals. If the top five journals are the same five journals that the seed journal cited the most, then these five journals are our core journals of analysis; otherwise, repeat the above two steps until a stable set of five journals is reached. This approach of determining core journals was pioneered by Hirst (Hirst, 1978) and was employed in our prior work (Yan & Zhu, 2015). The resulted five journals for each discipline are:

- **Astronomy & Astrophysics:** Astrophysical Journal**,** Monthly Notices of the Royal Astronomical Society**,** Astronomy & Astrophysics, Astrophysical Journal Letters, and Astronomical Journal
- **Computer Science, Theory & Methods**: Journal of the ACM, Theoretical Computer Science, Siam Journal on Computing, Information and Computation, and Journal of Computer and System Sciences
- **Engineering, Electrical & Electronic**: IEEE Transactions on Information Theory, IEEE Transactions on Wireless Communications, IEEE Transactions on Signal Processing, IEEE Transactions on Communications, and IEEE Journal on Selected Areas in Communications
- **Environmental Studies**[2]: Climatic Change**,** Global Environmental Change-Human and Policy Dimensions**,** Nature Climate Change**,** Journal of Climate**,** and Global Change Biology
- **Mathematics**: Inventiones Mathematicae, Annals of Mathematics, Duke Mathematical Journal, Lecture Notes in Mathematics, and Journal of the American Mathematical Society
- **Medicine, General & Internal**: New England Journal of Medicine**,** Lancet**,** JAMA-Journal of the American Medical Association**,** Annals of Internal Medicine**,** and BMJ-British Medical Journal
- **Nanoscience & Nanotechnology**: Nature Nanotechnology**,** Advanced Materials**,** Nano Today**,** Nano Letters**,** and ACS Nano

[2] Hirst's selection method did not converge for Environmental Studies because of the interdisciplinary nature of this domain. Instead, we manually picked five journals based on the citation relations of *Nature Climate Change* and *Global Environmental Change*.

The funding information can be gathered through the funding agency field (FU) in the output file in the Web of Science database for papers published after 2008 (Costas & Leeuwen, 2012; Paul-Hus, Desrochers, & Costas, 2016). The information in this field contains funders and grant IDs in brackets, with multiple funding sources delimited through semi-colons: for instance, "*National Science Foundation [CBET-1403871, DMR-1121107, CMMI-1150682, DGE-0946818]; National Science Foundation Graduate Research Fellowship [DGE-0946818]*". A few pilot studies on the use of Web of Science funding acknowledgement data have suggested that the funding agency field provides high precision and recall (both above 0.9) (Grassano, Rotolo, Hutton, Lang, & Hopkins, 2017). The limitation of the data set and the use of funding acknowledgment data is discussed in the *Limitations* section.

Bibliographic information of articles and review articles for each journal between January 2010 and March 2016 was downloaded. The total number of papers is 104,208; the disciplinary distribution is shown in Table 1.

Table 1. Data statistics

	No. of publications	Average no. of citations	Average no. of funding sources	Percentage of funding support
Astrophysics	50,392	13.99	4.10	89.22%
Computer Science	4,362	2.54	1.57	64.74%
Engineering	13,030	8.39	1.77	76.22%
Environmental Studies	8,111	15.93	2.39	85.81%
Mathematics	2,491	5.20	0.92	43.03%
Medicine	6,578	89.78	4.18	84.02%
Nanotechnology	19,244	40.78	3.21	95.46%

Table 1 includes papers both with and without funding. It shows that disciplines vary with respect to citation impact per paper, number of funding, sources per paper, and percentage of papers that received funding support. While the citation impact per paper in Medicine approaches 90, the average citation impact for three other disciplines is below 10: Engineering (8.39), Mathematics (5.20), and Computer Science (2.52). In terms of funding support, Medicine and Astrophysics are the disciplines with the highest *funding intensity*: each paper on average was supported by at least four different grants; meanwhile, Nanotechnology and Astrophysics are the disciplines with the highest *funding coverage*: about 90% of the papers in the two fields received some form of funding support. Mathematics, on the other hand, was the lowest in both funding intensity and funding coverage. We use Figure 1 to show the citation impact of papers with and without funding support among the seven domains.

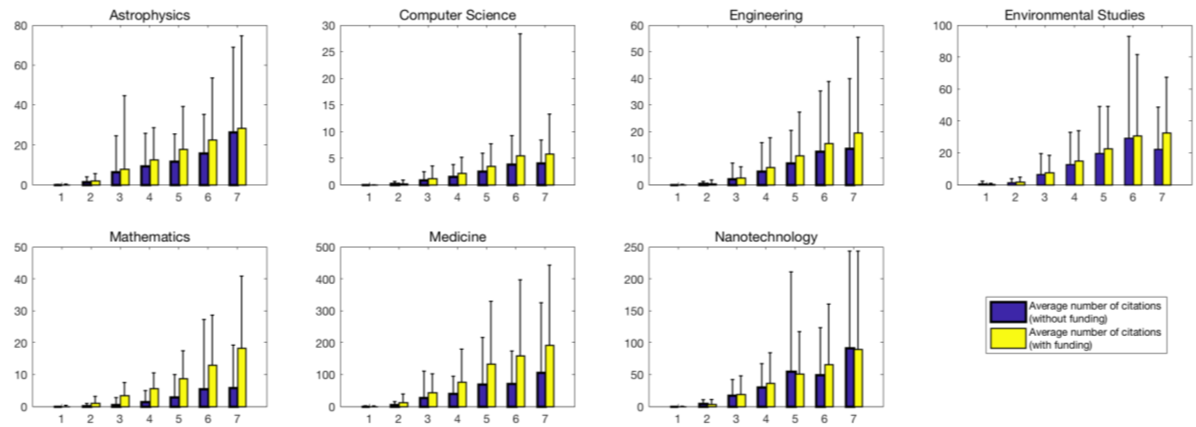

Figure 1. Average citation for papers with and without funding support

Figure 1 shows that, with the exception of Nanotechnology, for each number of year after publication, the citation impact of papers that received funding in STEMM fields is higher than those without funding support. This pattern is especially evident in Mathematics (8.37 citations per paper with funding support vs. 2.81 citations without funding support) and Medicine (97.20 citations per paper with funding support vs. 50.74 citations without funding support). For Nanotechnology, papers with funding support published in certain years received more citation impact than those without funding support (2011, 2013, 2014, and 2016), but pertained to fewer citations in other years. These results provide us with some visual cues on the association between citation impact and funding support. In the next section, we use a regression model to test the significance of the association.

Regression models

A linear regression model with Heckman bias correction is employed as the method of this study. A variety of indicators can be used to capture research outputs. Generally speaking, these indicators can be assigned to three categories (Boyack & Börner, 2003): activity measures (e.g., number of publications), impact measures (e.g., number of citations), and linkage measures (e.g., word co-occurrences). Among these, at the document level, impact measures best approximate research outputs and are the most widely used, since citations characterize formal scholarly communication and form the foundation of the modern scientific reward system (Cronin, 1984; Merton, 1968). In our regression analysis, we therefore employ the number of citations each paper received as the dependent variable. It is generally accepted in the research evaluation community that raw citation scores are not comparable across different disciplines; scholarly communication patterns, along with such proxies as average number of cited references and average number of coauthors per paper, simply vary too widely to make such comparisons informative (Hicks, Wouters, Waltman, De Rijcke, & Rafols, 2015; Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). The regression model used in this research thus treats disciplines as a categorical variable; within each discipline, we search for a demonstrable relationship between a paper's citation impact and its number of funding sources or number of coauthors. The regression model does not directly compare the citation scores or any other independent variables across disciplines; consequently, it may be seen as a collection of individual models—one for each discipline.

The Web of Science bibliographic data provide rich information on several key aspects of research, from which we extracted the following independent variables for each paper:

- *Funding*: the existence or absence of funding (for the models represented in Tables 2, 4, 5 and Table S1) or the number of funding sources that a paper reports in the acknowledgement or related applicable fields (for the model represented in Table 3);
- *Authors*: the number of authors a paper has;

- *Countries*: the number of countries that a paper's institutions are located in (we did not include the number of institutions as another variable to reduce the effect of multicollinearity);
- *References*: the number of papers that a paper cited;
- *Pages*: the length, in pages, of a paper; and
- *Years*: the number of years available for a paper to be cited. (A paper published in 2015, for example, has a *Years* value of 2).

These independent variables were selected because previous funder- or country-specific studies have shown them to be the most pertinent to a paper's citation impact (Gök et al., 2016).

We proceed to estimate the association between funding and papers' citations. We use the following linear regression model to study the log of number of citations:

$$Y_{it} = \eta_t + \beta X_{it} + \epsilon_{it}, \tag{1}$$

where η_t is the time fixed effect and X_{it} is a vector of independent variables. To properly estimate the association between funding and citations, we must account for the possible selection bias of funding. The fundamental logic is that certain unobservable characteristics may mark a project as higher-quality, making it both more likely to secure funding and more widely cited after publication. This selection bias would then lead to an overestimation of the funding factor. To compensate for this, we use the Heckman bias correction method (Heckman, 1976, 1979) to adjust for selection biases caused by certain unobservable characteristics such as the inherent quality of the articles, originality of the articles, and author reputation. Heckman bias correction is widely used for correcting selection bias in social sciences such as economics (Heckman, 1979), managerial (Certo, Busenbark, Woo, & Semadeni, 2016), and criminology (Bushway, Johnson, & Slocum, 2007). To apply it here, we use a probit model in the first stage to estimate the probability of an article getting funded:

$$P(funding|X_{it}) = \xi_t + \gamma X_{it}. \tag{2}$$

Equation (2) is the choice equation that estimates whether an article gets funded. Then we compute the inverse Mills ratio lambda

$$\lambda_{it} = \phi(\xi_t + \gamma X_{it})/\Phi(\xi_t + \gamma X_{it}), \tag{3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and cumulative distribution function of the standard normal distribution. Last, the inverse Mills ratio λ is added into Equation (1) and ordinary least squares are applied to estimate the parameters. For a more detailed discussion of Heckman bias correction, we refer readers to Wooldridge's popular text on the subject (Wooldridge, 2015).

Limitations

This research has two major limitations. First, this study did not distinguish the types of funding sources and, instead, they were treated as either a binary value or a count value. This treatment may confound the results because different types of funding supports should have varied effect on citation impact: for instance, a training or seed grant may have different outcome on citation impact from a multi-year research grant. Second, the employed regression model is capable of capturing the associations between the independent variables and citation impact, but it could not verify the causality of the independent variables such as funding that lead to the gain in citation impact. The above Heckman selection models, despite some criticisms in light of recent developments in econometrics research, are still commonly used in social sciences. In case of causal inference is preferred over the association analysis, some *exogenous shocks* can be exploited. A shock generates "natural" or "quasi" experiments, which can explain that a change in government funding caused the behaviors of research outputs. For example, an exogenous policy change in science funding in the new administration may provide opportunities to gauge the impact of such policy (or the lack of it) (Park, Lee, & Kim, 2015). We see shock-base experiments as a future research direction.

Results

Table 2 reports the observed extent of the association between funding and citations, after Heckman bias correction has been applied. In our research setting, there may exist selection biases because funding agencies do not select research projects randomly. It may well be the case that higher quality projects were selected by funding agencies and meanwhile higher quality projects may also produce papers with higher citations. Thus, the role of funding on citation may be overestimated without considering the selectivity of funding. Columns 1-7 report the estimates for STEMM disciplines. The inverse Mills ratio term *lambda* is significant and negatively signed. This suggests that the unobserved factors (e.g., quality of the paper), though they boost the likelihood of attracting funding, tend to be associated with fewer citations. Note that the estimated coefficient of *funding* is slightly smaller than the ordinary least squares (OLS) estimate (Table S1), indicating that the presence of funding gave an upward bias to the number of citations.

Table 2. Parameter estimation after Heckman bias correction.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(citations)	Astro	CS	Eng	Env	Math	Med	Nano
Funding	0.164***	0.072***	0.146***	0.154***	0.813***	0.565***	0.129***
	(0.013)	(0.023)	(0.019)	(0.028)	(0.036)	(0.041)	(0.035)
# Authors	0.003***	0.021	0.043***	0.028***	-0.124***	0.001	0.025***
	(0.000)	(0.017)	(0.012)	(0.004)	(0.028)	(0.001)	(0.006)
Pages	0.012***	0.001	0.019***	-0.027***	0.003***	0.012**	-0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.005)	(0.003)
# Countries	-0.026***	-0.029	0.020	0.023**	-0.524***	0.036***	-0.029***
	(0.006)	(0.026)	(0.016)	(0.010)	(0.066)	(0.005)	(0.010)
# References	0.004***	0.007***	0.009***	0.004***	-0.006***	0.006***	0.007***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.000)
Years	0.438***	0.273***	0.451***	0.516***	0.303***	0.584***	0.591***
	(0.002)	(0.006)	(0.005)	(0.006)	(0.010)	(0.008)	(0.004)
lambda	-3.200***	-0.603**	-0.384**	-1.782***	-3.037***	-2.609***	-0.946**
	(0.218)	(0.246)	(0.181)	(0.222)	(0.301)	(0.113)	(0.474)
Constant	0.319***	-0.292	-1.137***	0.263***	3.155***	0.708***	-0.111
	(0.074)	(0.242)	(0.143)	(0.088)	(0.439)	(0.076)	(0.092)
Observations	50,392	4,362	13,030	8,111	2,491	6,578	19,244
R-squared	0.467	0.340	0.409	0.511	0.500	0.486	0.521

Standard errors of the regression estimates are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2 shows that except for Mathematics that funding is the most significant indicator, years after publication is the most significant indicator of a paper's citation impact for all other domains. Variables including number of authors, number of pages, number of countries, and number of references only have marginal effect on citations for all other domains but Mathematics. In Mathematics, the number of countries and number of authors have a negative association with citation impact: the coefficients translate into 41% and 12% citation deductions respectively ($e^{-0.124}$ and $e^{-0.524}$). Positive coefficients of *Funding* for STEMM disciplines indicate that funded papers do tend to have more citations. The magnitude of funding seems to be most pronounced in Mathematics, where the coefficient of 0.813 translates into 125% more citations for funded papers than non-funded papers ($e^{0.813}$). It is least pronounced in Computer Science publications: the coefficient of 0.072 translates to just 7% more citations for funded papers than non-funded papers.

We also report the relationship between number of funding sources and number of citations in Table 3. As with Table 2, these results are obtained after correcting for the selective nature of funded projects.

Table 3. Parameter estimation after Heckman bias correction for number of funding agencies

Log(citation)	(1) Astro	(2) CS	(3) Eng	(4) Env	(5) Math	(6) Med	(7) Nano
# Agencies	0.005***	0.022***	0.034***	0.035***	0.197***	0.016***	0.039***
	(0.001)	(0.006)	(0.005)	(0.005)	(0.014)	(0.002)	(0.004)
# Authors	0.003***	0.021	0.040***	0.028***	-0.180***	0.000	0.024***
	(0.000)	(0.017)	(0.012)	(0.004)	(0.029)	(0.001)	(0.006)
Pages	0.012***	0.001	0.020***	-0.027***	0.004***	0.020***	-0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.005)	(0.003)
# Countries	-0.031***	-0.034	0.010	0.014	-0.724***	0.029***	-0.046***
	(0.006)	(0.026)	(0.016)	(0.010)	(0.069)	(0.005)	(0.010)
# References	0.004***	0.007***	0.009***	0.004***	-0.009***	0.006***	0.006***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.000)
Years	0.438***	0.273***	0.451***	0.516***	0.305***	0.581***	0.592***
	(0.002)	(0.006)	(0.005)	(0.006)	(0.010)	(0.008)	(0.004)
lambda	-3.386***	-0.560**	-0.351*	-1.667***	-3.909***	-2.881***	-0.510
	(0.218)	(0.247)	(0.182)	(0.223)	(0.313)	(0.112)	(0.474)
Constant	0.510***	-0.297	-1.082***	0.294***	4.580***	1.168***	-0.122
	(0.072)	(0.242)	(0.142)	(0.085)	(0.454)	(0.068)	(0.085)
Observations	50,392	4,362	13,030	8,111	2,491	6,578	19,244

R-squared	0.466	0.340	0.409	0.512	0.444	0.476	0.524

Standard errors of the regression estimates are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

While citations still tend to be higher for papers with more funding supports in STEMM disciplines when the number of funding sources is considered, the difference of citation impact between funded and non-funded papers is much narrower. The funding advantage ranges from 0.5% increase for Astrophysics to 22% increase for Mathematics. The average citation increase is only 5% for funded papers compared with non-funded ones, which is significantly lower than the average reported in Table 2 (39%) when funding supported is treated as a binary variable.

It is also interesting to investigate how authors would synergize with funding. For example, is funding more salient if the article is multi-authored? Does co-authorship magnify the association between funding and citations? If there is such a synergistic relationship, does it apply differently to STEMM disciplines? We show the results in Table 4 where funding is binary variable equal to 1 if the paper is funded and 0 otherwise; Multiple Authors is an indicator variable that equal to 1 if the paper has two or more authors and 0 otherwise; and Funding*Multiple Authors is the interaction term.

Table 4. Parameter estimation after Heckman bias correction with coauthor and funding as an interaction factor

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(citation)	Astro	CS	Eng	Env	Math	Med	Nano
Funding	0.278***	0.079	0.130**	0.363***	0.769***	0.239*	0.445**
	(0.035)	(0.056)	(0.066)	(0.080)	(0.056)	(0.141)	(0.211)
Multiple Authors	0.219***	0.070	0.141***	0.285***	-0.054	-0.433***	0.181
	(0.031)	(0.046)	(0.053)	(0.068)	(0.055)	(0.106)	(0.176)
*Funding*Multiple Authors*	0.356***	0.136***	0.283***	0.388***	0.825***	0.178*	0.298*
	(0.029)	(0.045)	(0.052)	(0.063)	(0.056)	(0.102)	(0.173)
Pages	0.012***	0.000	0.019***	-0.026***	0.003***	0.013***	-0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.005)	(0.003)
# Countries	0.024***	-0.036	0.011	0.069***	-0.441***	0.036***	-0.024**
	(0.004)	(0.023)	(0.016)	(0.008)	(0.065)	(0.005)	(0.010)
# References	0.005***	0.007***	0.008***	0.004***	-0.002	0.006***	0.007***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Years	0.439***	0.272***	0.451***	0.519***	0.302***	0.584***	0.591***
	(0.002)	(0.006)	(0.005)	(0.006)	(0.010)	(0.008)	(0.004)
lambda	-1.630***	-0.733***	-0.728***	-2.626***	-2.212***	-2.703***	-2.820***

	(0.192)	(0.148)	(0.115)	(0.191)	(0.265)	(0.114)	(0.207)
Constant	-0.390***	-0.188	-0.945***	0.311***	1.952***	1.108***	0.050
	(0.071)	(0.150)	(0.104)	(0.101)	(0.380)	(0.122)	(0.178)
Observations	50,392	4,362	13,030	8,111	2,491	6,578	19,244
R-squared	0.466	0.340	0.410	0.509	0.497	0.487	0.521

Standard errors of the regression estimates are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

As the results in Table 4 demonstrate, such a synergy does exist: for most disciplines, the interaction term of *Funding* and *Multiple Authors* has a significant coefficient. Positive coefficients for *Funding*Multiple Authors* in STEMM disciplines suggest that funded, multi-author papers tend to have even more citations than funded, single-author papers. In other words, multiple authorship indeed tends to magnify the association between funding and citations observed in our earlier results. In the case of Mathematics, the number of coauthors alone is not associated with gain in citations (which is evidenced by the negative coefficients), but when a coauthored paper received funding support, a significant gain in citation is achieved. With multiple authorships, the positive association of funding is the largest in Mathematics, where the coefficient of 0.825 translates into 128% more citations for funded papers than non-funded papers, followed by Environmental Studies with 47% more citations and Astrophysics with 43% more citations. It is least pronounced in Computer Science and Medicine: the coefficient of 0.136 translates to just 15% more citations for Computer Science and the coefficient of 0.178 translates to 19% for Medicine.

Finally, we investigate how a paper's institutional affiliations synergize with funding. Table 5 reports the interaction between funding and multiple institutions where multiple Institutions is an indicator variable that equal to 1 if the paper has two or more institutions and 0 otherwise and Funding*Multiple Institutions is the interaction term.

Table 5. Parameter estimation after Heckman bias correction with multi-institutional affiliation and funding as an interaction factor

Log(citation)	(1) Astro	(2) CS	(3) Eng	(4) Env	(5) Math	(6) Med	(7) Nano
Funding	0.287***	0.103*	0.171**	0.315***	0.782***	0.272*	0.195
	(0.038)	(0.058)	(0.067)	(0.082)	(0.058)	(0.162)	(0.228)
Multiple Institutions	0.228***	0.110**	0.146***	0.244***	-0.029	-0.428***	0.107

	(0.034)	(0.048)	(0.054)	(0.071)	(0.055)	(0.132)	(0.193)
*Funding*Multiple Institutions*	0.369***	0.171***	0.285***	0.360***	0.836***	0.160	0.229
	(0.032)	(0.047)	(0.053)	(0.066)	(0.058)	(0.130)	(0.191)
Pages	0.012***	0.000	0.019***	-0.026***	0.003***	0.013***	-0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.005)	(0.003)
# Countries	0.024***	-0.033	0.010	0.069***	-0.435***	0.036***	-0.024**
	(0.004)	(0.023)	(0.016)	(0.008)	(0.065)	(0.005)	(0.010)
# References	0.005***	0.007***	0.008***	0.004***	-0.002	0.006***	0.007***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Years	0.439***	0.272***	0.451***	0.519***	0.302***	0.584***	0.591***
	(0.002)	(0.006)	(0.005)	(0.006)	(0.010)	(0.008)	(0.004)
lambda	-1.630***	-0.688***	-0.747***	-2.616***	-2.166***	-2.698***	-2.790***
	(0.192)	(0.148)	(0.116)	(0.193)	(0.269)	(0.114)	(0.206)
Constant	-0.404***	-0.256*	-0.939***	0.337***	1.883***	1.130***	0.116
	(0.072)	(0.150)	(0.105)	(0.104)	(0.386)	(0.147)	(0.196)
Observations	50,392	4,362	13,030	8,111	2,491	6,578	19,244
R-squared	0.466	0.340	0.409	0.508	0.496	0.487	0.521

Standard errors of the regression estimates are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The results are qualitatively similar to the interaction between funding and multiple authorships in Table 4. That is, multiple institutions magnify the funding association yet again: in STEMM disciplines, funded multi-institution papers tend to obtain even more citations than funded single-institution papers. With multiple institutions, the positive association of funding is the largest in Mathematics, where funded papers received 130% more citations than non-funded papers, followed by Astrophysics with 45% more citations and Environmental Studies with 43% more citations. It is least pronounced in Medicine and Computer Science: funded Medicine papers received 17% more citations than non-funded papers and funded Computer Science papers received 19% more citations than non-funded papers.

Discussion and conclusion

The results obtained in this study are consistent with previous agency- and domain-specific studies in that a significant, though sometimes weak, relationship was found between citation impact and funding (Fortin & Currie, 2013; Jacob & Lefgren, 2011; Rigby, 2013). More importantly, this study extended the scope of prior research by examining disciplinary differences in funding reception; we found that, in terms of citation impact, STEMM fields tend to respond more favorably to funding, particularly for

Mathematics. While funded papers in all other domains had more gain in citation impact when multi-authorship and multi-institutional affiliation were considered as interaction terms with funding, funded papers in Medicine received less gain in citation impact when either term was considered, suggesting that Medicine has a rather different funding mechanism from STEM domains. The identified synergistic properties of multiple authorship and multi-institutional affiliation are another novel finding.

The number of years after publication is no doubt the most significant indicator of a paper's citations. However, funding (considered as a binary variable) is the second most significant indicator in estimating papers' citations and is stronger than any other variables included in the model. In STEMM disciplines, funding helps investigators form large collaborative teams who are more likely to produce and publish high-impact research (Borsuk, Budden, Leimu, Aarssen, & Lortie, 2009; Figg et al., 2006; Franceschet & Costantini, 2010). When we measured funding in terms of number of funding supports, the association on the output of citations is less strong; this suggests that while having a funding support is markedly associated with a paper's citation impact, the number of citations tends not to respond strongly as the number of funding supports increases. The policy implication of this finding is that, from the citation impact point of view, it may be more favorable to allocate funding to a wider range of researchers, particularly those without active funding support, instead of awarding researchers who already have multiple funding sources at hand. The tendency of resources to concentrate in the hands of a few—the so-called "Matthew effect" in science—has already been observed in prior research (Azoulay, Stuart, & Wang, 2013; Larivière & Gingras, 2010). With this study, we found that such a pattern of funding allocation is not conducive to boost the citation impact of papers that could otherwise be supported by funding.

Future research will involve interlinking Web of Science publication metadata with STAR METRICS via grant numbers, which will allow for more granular analyses of publications funded by U.S. federal agencies. It will also identify agencies from the funding metadata and assess the role that certain funding agencies (e.g., NSF, NIH, and DARPA) have in promoting publications' citation impact. Another direction is to conduct causal inference on government funding. Events such as the COMPETES Act passed by House in May 2015 (Mervis, 2015) and the EPA funding freeze by the new administration (Dennis & Eilperin, 2017) can serve as exogenous shocks and allow us to draw causal inferences for the policy impact in terms of research outputs (i.e., publications and citations).

Acknowledgements

This project was made possible in part by the Institute of Museum and Library Services (Grant Award Number: RE-07-15-0060-15), for the project titled "Building an entity-based research framework to enhance digital services on knowledge discovery and delivery". This work was also partly supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A3A2046711).

References

Álvarez-Bornstein, B., Morillo, F., & Bordons, M. (2017). Funding acknowledgments in the Web of Science: completeness and accuracy of collected data. *Scientometrics, 112*(3), 1793-1812.

Azoulay, P., Stuart, T., & Wang, Y. (2013). Matthew: Effect or fable? *Management Science, 60*(1), 92-109.

Borsuk, R., Budden, A., Leimu, R., Aarssen, L., & Lortie, C. (2009). The influence of author gender, national language and number of authors on citation rate in ecology. *Open Ecology Journal, 2*, 25-28.

Boyack, K. W., & Börner, K. (2003). Indicator‐assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology, 54*(5), 447-461.

Boyack, K. W., & Klavans, R. (2015). *Is the most innovative research being funded.* Paper presented at the 20th International Conference on Science and Technology Indicators.

Bushway, S., Johnson, B. D., & Slocum, L. A. (2007). Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology, 23*(2), 151-178.

Campbell, D., Picard-Aitken, M., Côté, G., Caruso, J., Valentim, R., Edmonds, S., . . . Bastien, N. (2010). Bibliometrics as a performance measurement tool for research evaluation: The case of research funded by the National Cancer Institute of Canada. *American Journal of Evaluation, 31*(1), 66-83.

Certo, S. T., Busenbark, J. R., Woo, H. s., & Semadeni, M. (2016). Sample selection bias and Heckman models in strategic management research. *Strategic management journal, 37*(13), 2639-2657.

Costas, R., & Leeuwen, T. N. (2012). Approaching the "reward triangle": General analysis of the presence of funding acknowledgments and "peer interactive communication" in scientific publications. *Journal of the American Society for Information Science and Technology, 63*(8), 1647-1661.

Cragin, M. H., Nichols, L., Simon, M., & Watts, S. M. (2012). Measuring science: Emerging tools for analysis of federal R&D investments. *Proc. ASIST Annu. Meet, 49*(1).

Cronin, B. (1984). *The citation process. The role and significance of citations in scientific communication*. London: Taylor Graham.

Cronin, B., & Shaw, D. (1999). Citation, funding acknowledgement and author nationality relationships in four information science journals. *Journal of Documentation, 55*(4), 402-408.

Dennis, B., & Eilperin, J. (2017). Trump administration tells EPA to freeze all grants, contracts. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/energy-environment/wp/2017/01/23/trump-administration-tells-epa-to-freeze-all-grants-contracts/?utm_term=.1c3f8ef75a0d

Figg, W. D., Dunn, L., Liewehr, D. J., Steinberg, S. M., Thurman, P. W., Barrett, J. C., & Birkinshaw, J. (2006). Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 26*(6), 759-767.

Fortin, J.-M., & Currie, D. J. (2013). Big science vs. little science: how scientific impact scales with funding. *PloS one, 8*(6), e65263.

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics, 4*(4), 540-553.

Gök, A., Rigby, J., & Shapira, P. (2016). The impact of research funding on scientific outputs: Evidence from six smaller European countries. *Journal of the Association for Information Science and Technology, 67*(3), 715-730.

Grassano, N., Rotolo, D., Hutton, J., Lang, F., & Hopkins, M. M. (2017). Funding data from publication acknowledgments: Coverage, uses, and limitations. *Journal of the Association for Information Science and Technology, 68*(4), 999-1017.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models *Annals of Economic and Social Measurement* (Vol. 5, pp. 475-492): NBER.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica: Journal of the Econometric Society*, 153-161.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429-431.

Hirst, G. (1978). Discipline impact factors: A method for determining core journal lists. *Journal of the American Society for Information Science, 29*(4), 171-172.

Holton, G. (1978). Can science be measured *Toward a Met ric of Science: The Advent of Science Indicators*. New York: Social Science Research Council.

Jacob, B. A., & Lefgren, L. (2011). The impact of research grant funding on scientific productivity. *Journal of public economics, 95*(9), 1168-1177.

Lane, J. (2009). Assessing the impact of science funding. *Science, 324*(5932), 1273-1275.

Lane, J., & Bertuzzi, S. (2010). *The STAR METRICS project: current and future uses for S&E workforce data.* Paper presented at the Science of Science Measurement Workshop, Held in Washington DC.

Lane, J., & Bertuzzi, S. (2011). Measuring the results of science investments. *Science, 331*(6018), 678-680.

Lane, J., Owen-Smith, J., Rosen, R. F., & Weinberg, B. A. (2015). New linked data on research investments: Scientific workforce, productivity, and public value. *Research policy, 44*(9), 1659-1671.

Largent, M. A., & Lane, J. I. (2012). STAR METRICS and the science of science policy. *Review of Policy Research, 29*(3), 431-438.

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology, 61*(2), 424-427.

Leydesdorff, L., & Wagner, C. (2009). Macro-level indicators of the relations between research funding and research output. *Journal of Informetrics, 3*(4), 353-362.

Ma, A., Mondragón, R. J., & Latora, V. (2015). Anatomy of funded research in science. *Proceedings of the National Academy of Sciences, 112*(48), 14760-14765.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56-63.

Mervis, J. (2015). After 2-year battle, House passes COMPETES Act on mostly party-line vote. *Science*.

National Science Board. (2016). Science and Engineering Indicators 2016.
Park, H., Lee, J. J., & Kim, B.-C. (2015). Project selection in NIH: A natural experiment from ARRA. *Research policy, 44*(6), 1145-1159.
Paul-Hus, A., Desrochers, N., & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics, 108*(1), 167-182.
Rigby, J. (2013). Looking for the impact of peer review: does count of funding acknowledgements really predict research impact? *Scientometrics, 94*(1), 57-73.
Sarli, C. C., & Carpenter, C. R. (2014). Measuring academic productivity and changing definitions of scientific impact. *Missouri medicine, 111*(5), 399.
Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: Queries and Caveats. *Journal of the Association for Information Science and Technology, 68*(3), 790-794.
Trochim, W. M., Marcus, S. E., Mâsse, L. C., Moser, R. P., & Weld, P. C. (2008). The Evaluation of Large Research Initiatives A Participatory Integrative Mixed-Methods Approach. *American Journal of Evaluation, 29*(1), 8-28.
Van Noorden, R. (2015). Seven thousand stories capture impact of science. *Nature, 518*(7538), 150.
Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37-47.
Wang, J., & Shapira, P. (2011). Funding acknowledgement analysis: an enhanced tool to investigate research sponsorship impacts: the case of nanotechnology. *Scientometrics, 87*(3), 563-586.
Wang, X., Liu, D., Ding, K., & Wang, X. (2011). Science funding and research output: a study on 10 countries. *Scientometrics, 91*(2), 591-599.
Weinberg, B. A., Owen-Smith, J., Rosen, R. F., Schwarz, L., Allen, B. M., Weiss, R. E., & Lane, J. (2014). Science funding and short-term economic activity. *Science, 344*(6179), 41-43.
Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*: Nelson Education.
Yan, E., & Zhu, Y. (2015). Identifying entities from scientific publications: A comparison of vocabulary- and model-based methods. *Journal of Informetrics, 9*(3), 455-465.
Yong, E. (2017). Thanks to Trump, Scientists Are Planning to Run For Office. *The Atlantic*. Retrieved from https://www.theatlantic.com/science/archive/2017/01/thanks-to-trump-scientists-are-planning-to-run-for-office/514229/
Zhao, D. (2010). Characteristics and impact of grant-funded research: a case study of the library and information science field. *Scientometrics, 84*(2), 293-306.