CESifo WORKING PAPERS

10661
2023

September 2023

Do Twitter's Science Stars Get a Citation Premium?

Christian Lessmann, Ali Sina Önder

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10661*

Do Twitter's Science Stars Get a Citation Premium?

Abstract

We analyze whether the social media popularity of Twitter star scientists, who were identified by Science in a 2014 report, pays off in terms of an increased number of citations. To establish a causal relationship, we use the COVID-19 global pandemic as a quasi-natural experiment exogenously increasing public attention and the demand for expertise. Using Twitter science stars' and their coauthors' publications on COVID related topics prior to the break out of the pandemic, we run a difference-in-differences analysis for annual incoming citations of the two groups. We find that the Twitter star status added about 1.07 extra citations following the breakout of COVID-19 per year per article, corresponding to about 70% of the already existing citation gap between Twitter science stars and their coauthors. Moreover, we also document that the publication of the Science list on Twitter science stars in 2014 per se caused an increase in citations, i.e. the publication of the supposed celebrity status by Science already benefited the stars, which meant 1.06 more citations per year per article compared to their coauthors. Treatment based on scientists' Kardashian indexes yields no robust effects, implying that unjustified social media popularity does not pay off in terms of citations.

JEL-Codes: J240, O330.

Keywords: social media, expertise, Kardashian index, citations, Covid.

Christian Lessmann
Dresden University of Technology
Dresden / Germany
christian.lessmann@tu-dresden.de

*Ali Sina Önder**
University of Portsmouth
Portsmouth / United Kingdom
ali.onder@port.ac.uk

*corresponding author

September 13, 2023
We are grateful to Scott Mahadeo and Sascha Schweitzer for their insightful comments and discussions. We thank Aslak Butler for research assistance. We are especially grateful to Sahar Abuelbashar and Katie Osgood for their help with SciVal and Scopus. The usual disclaimer applies.

1 Introduction

In 2014, *Science* released a list of top 50 resp. 100 'Twitter science stars' or 'Twitter's scientific celebrities'[1] (Travis, 2014, You, 2014) in an attempt to shed light on then newly introduced concept of Kardashian index (Hall, 2014). According to Hall (2014), there is a level of social media followers that is justified by the scientific prominence of a researcher, which is captured by researcher's citation accumulation. If the number of followers exceeds this justified level, this contributes to the researcher's *Kardashian*ness. You (2014) and Travis (2014) rank scientists from different disciplines according to their number of followers on Twitter, adding information on their citation count and their Kardashian index. In addition to its relation to the discussion around the then newly introduced Kardashian index, the list may be seen as a pure fun fact on science communication. However, it is also useful to analyze the perceived degree of expertise and popularity among Twitter science stars and the impact of their social media activity on academic reputation, i.e., citations.[2]

Using the Travis (2014) list, we analyze whether the social media popularity of Twitter science stars pays off in terms of an increased number of citations. To establish a causal relationship, we use the COVID-19 global pandemic as a quasi-natural experiment exogenously increasing public attention and the demand for expertise. In our empirical analysis, we concentrate on biology and virology (esp. immunology, pharmacology, biochemistry as explained in Section 2) among the Twitter science stars who published on COVID-related topics prior to the break out of the pandemic.

We research their relevant publications and corresponding incoming citations and compare citations to those of their co-authors on similar topics. Papers by Twitter science stars are in the treatment group (treatment activates in 2019), while papers of coauthors form the control group, whereby we exclude joint work with the stars. We find that the Twitter star status added about 1.09 extra citations following the breakout of the COVID-19 per year per article to the already existing citation gap between the Twitter science stars and their coauthors. The citation gap widens by about 70% of the pre-pandemic level. Moreover, we

[1] We are aware of the fact that Elon Musk renamed the platform in 2023 to $\mathbb{X}$, but we keep the old and popular name Twitter for simplicity.

[2] We refer to the scientists with great presence on Twitter as collected by Travis (2014) as Twitter science stars throughout this paper.

also document that the publication of the *Science* list on Twitter science stars in 2014 per se caused an increase in citations, i.e. the publication of the supposed celebrity status by *Science* already benefited the stars, which meant 1.06 more citations per year per articles compared to their coauthors. Concerning the Kardashian index, we find no robust effects implying that unjustified social media popularity does not pay off in terms of citations.

Our study adds to the literature on the rewards of social media activity for scientists. Chan et al. (2023) are among the first to establish a causal relationship between social media activity on a research paper and the number of citations received by that paper: Twitter engagement with an economics working paper leads up to 25% more citations when the paper is published. This is, however, a limited way to think about the impact of social media. As Hall (2014, p.1) stated, "in the age of social media there are people who have high-profile scientific blogs or Twitter feeds but have not actually published many peer-reviewed papers of significance; in essence, scientists who are seen as leaders in their field simply because of their notoriety." Our study shows that the notoriety mentioned by Hall (2014) does not pay off for itself. Being active on Twitter pushes citations, but the effect is not robust for researchers scoring high on the Kardashian index.

Our paper is organized as follows. Section 2 introduces data and methodology, Section 3 shows the results, and Section 4 concludes.

2 Data and Methodology

We gather annual inflow of citations to journal articles of Twitter science stars and their coauthors using the Scopus database. At this, we concentrate on those scientists from Travis (2014)'s list of top 100 Twitter science stars who have peer-reviewed journal publications in Scopus's subject areas of *immunology and microbiology* or *pharmacology, toxicology, and pharmaceutics* or *biochemistry, genetics, and molecular biology* between 2001 and 2015.[3] Ideally, we need to identify those who have published in related areas to virology, infectious diseases, esp. in SARS, MERS, or H1N1 (swine flu), which we refer to as *COVID-related*

[3]Usually an articles' annual citations peaks about three years, which is the reason we use 2015 as the cut-off year for our analysis. Citations' life cycle is important for the identification strategy that is discussed in this section.

research in this paper. A publication in the above-mentioned broader subject areas does not have to be a COVID-related research, thus, we checked scientists' publications to make sure that we include only those Twitter science stars who have *COVID-related* research. Next, we identified Twitter science stars' coauthors as follows: A coauthor is someone who appears as a coauthor on a Twitter science star's papers identified in COVID-related research areas and has a journal publication in these areas that is *not* coauthored with the respective Twitter science star between 2011 and 2015. This leaves us with 13 Twitter science stars and 20 coauthors. Names of stars and their respective coauthors are listed in Table A.1 in Appendix.

Our dataset contains a total of 1,304 journal articles of a total of 33 scientists. 13 Twitter science stars produced 761, and their 20 coauthors produced 543 of these articles. Twitter science stars' articles contain all of their coauthors but coauthors' articles do not contain any papers coauthored with the respective Twitter science star.[4] We collected annual citation inflow data from Scopus for each article in our data set. We use the academic age of scientists (calculated as years from their first ever publication) and the total number of publications in their career up to any given year as controls in our analysis.

Our identification strategy is based on the exogenous attention shock caused by COVID-19: Twitter science stars and their coauthors may have different amounts of annual citation inflow to their journal articles over the years but if their annual citation inflows have parallel trends before 2019 and we observe a significant change in the difference of annual citations for their pre-2019 COVID-related research from the treatment in 2019 on, then we show that Twitter science stars enjoy a premium of visibility. There is no obvious reason why Twitter science stars' pre-2019 COVID-related research should attract disproportionately more attention than that of their coauthors after 2019; any widening in their citation difference can be attributed to the high visibility premium that Twitter science stars enjoy due to their expert status on social media.

[4]Twitter science stars have many coauthors, but many of these coauthors either did not publish in the subject areas of our interest during 2011-2015 or they did not publish without the respective Twitter science star being a coauthor.

We estimate the following difference-in-differences (DiD) model:

$$Citations_{pst} = \alpha_1 Time + \alpha_2 Treated_s + \alpha_3 Time \times Treated_s + \beta_1 X_p + \beta_2 X_{st} + \phi_t + \phi_j + \phi_s + \epsilon_{pst} \tag{1}$$

Incoming citations (in logs) to paper p of scientist s in year t is regressed on time, treatment, their interactions, as well as publication characteristics (X_p), and time-variant scientist characteristics (X_{st}). We use year, journal, and individual scientist fixed effects (ϕ_t, ϕ_j, ϕ_s). We provide two sets of estimations where the time for treatment is active post-2014 in one and post-2019 in the other. These models capture the DiD effects of the publication of *Science*'s top Twitter scientists list that draws attention to the Twitter science stars and the perceived expertise of Twitter scientists on COVID-19, respectively. In either case, α_3 is the main coefficient of interest. As an alternative to being a Twitter science star, we use the Kardashian index (K-index) as the treatment. K-index is a continuous variable, and although it has low values for coauthors of Twitter science stars, it is non-zero for many and there is a wider variation among science stars as well.

3 Results

Panel (a) of Figure 1 shows means of annual citations of Twitter science stars' and coauthors' COVID-related research that was published between 2002 and 2015. We observe a boost after 2019: Papers of Twitter science stars who worked on COVID-related research that were published up to 2015 received a citation boost after 2019 compared to their coauthors' citations. COVID-19 was a worldwide pandemic that caught unprecedented attention from the media and public as it was a world-wide emergency situation, and expert opinion was highly valued at that time (Lavezzolo et al., 2022). The main argument put forward by Hall (2014) is that a researcher with high visibility in social media may get disproportionately more public attention. There is a widening of the difference in 2020 and coauthors seem to catch up with stars with one year delay although there is no plausible reason as to why Twitter science stars should get such a citation boost except that they were perceived as experts to a higher degree than their coauthors due to their visibility.

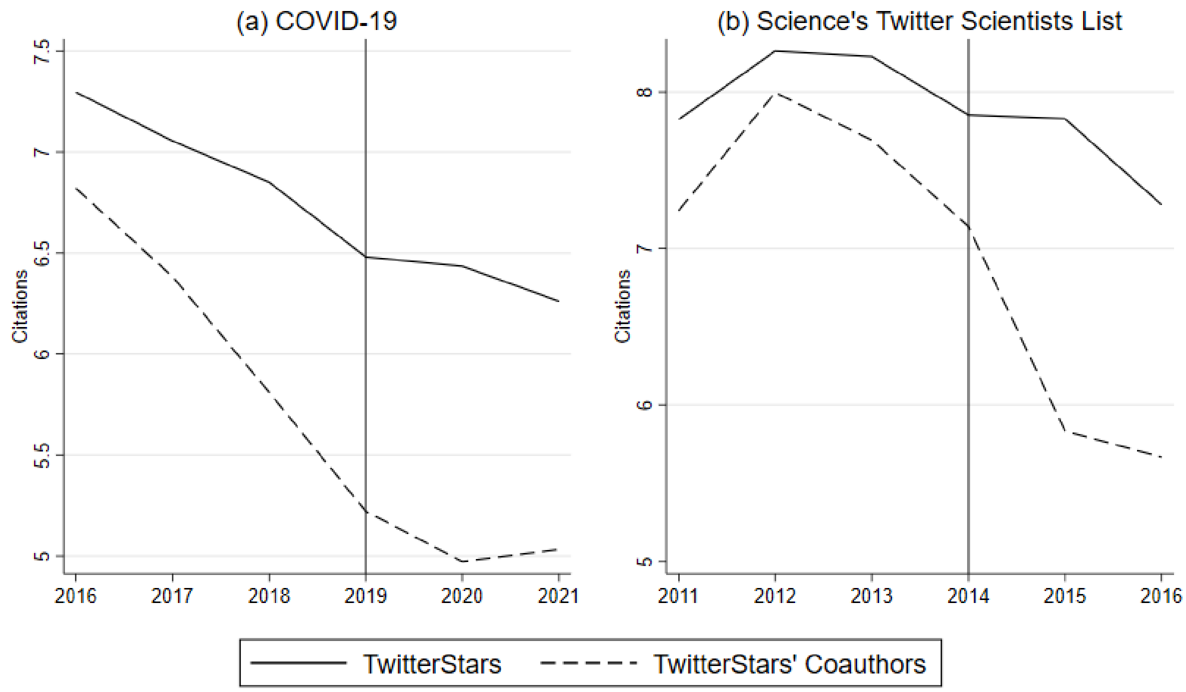

Figure 1: Means of annual citation inflow to journal articles of TwitterStars and their coauthors.

Times of the COVID-19 pandemic was not the first time we observe such a social media visibility effect. Panel (b) of Figure 1 shows the means of annual citations of the same two groups' articles that were published between 2002 and 2010. Twitter science stars enjoyed a boost to their annual citations after 2014 that were published before 2010. 2014 is the year when Science's Twitter science stars list was published. Following the publication of the list, we observe a large increase in the difference between annual citations received by Twitter stars' and their coauthors' research. The same argument as for the COVID-19 effect can be made for Science's list as well: There is no obvious reason other than the very publication of the Twitter science star list why Twitter science stars' research should get disproportionately more citations during this time.

Figure A.1 in the Appendix provides a brief summary of our test for the parallel trends assumption. Annual citation trends are shown to be parallel before treatments in either case were active. The null hypothesis of parallel trends is tested in difference-in-differences models for events of COVID-19 and Science's list including quality controls, age polynomials, and fixed effects for individual, year, journal, and publication year. For COVID-19, we obtain F-value=0.45 (p=0.51) and for Science's list event we obtain F-value=2.22 (p=0.15), hence the null hypothesis of parallel trends cannot be rejected in either event.

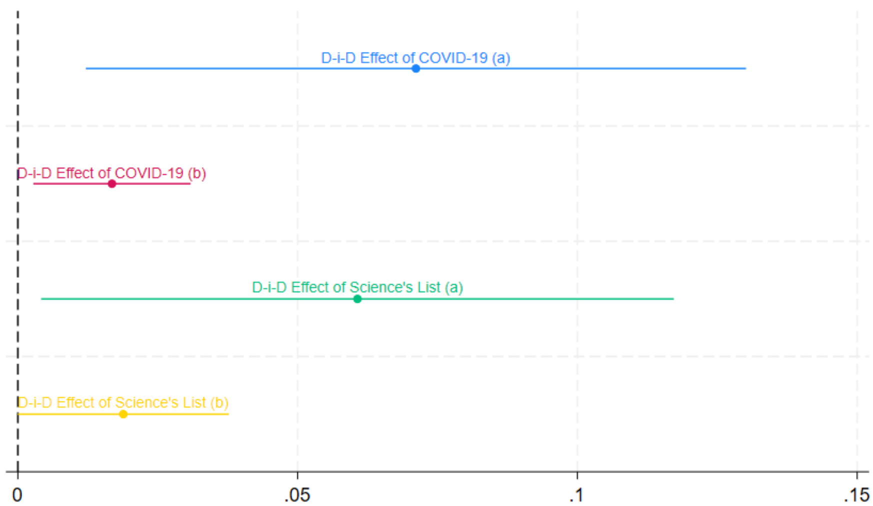

Figure 2: Difference-in-differences coefficient estimates with 90% confidence interval. All four models contain quality controls, age polynomials, individual, year, journal, and publication year fixed effects. Treatment in models labeled (a) is a binary variable whether a scientist was listed as a Twitter science star by Travis (2014) or not; treatment in models labeled (b) is the K-index of scientists.

We run DiD models for both the COVID-19 effect and the Science list effect using the specification of equation 1 where treatment is being a Twitter science star and K-index in models (a) and (b), respectively. Coefficient plots are shown in Figure 2. Although the DiD effect for COVID-19 has higher point estimates than that for the Science list, they are within the 90% confidence interval of each other, and more importantly, all four estimates are statistically significantly different than zero. Estimated coefficients and diagnostic statistics are shown in Tables A.2 and A.3 in the Appendix.[5]

Both treatment effects, being a Twitter science star and the K-index yield significant difference-in-differences. Coefficient estimate in column (2) of Table A.2 reveals that Twitter science star status added about 1.07 extra citations following the breakout of the COVID-19 pandemic per year per article to the already existing citation gap between the Twitter

[5] Coefficients depicted in Figure 2 correspond to the coefficient estimate of the interaction of treatment with post-2019 and post-2014 years in columns (2) and (4) of Tables A.2 and A.2, respectively.

scientist and their coauthors. Considering that the average annual citation difference between Twitter science stars and their coauthors was about 1.5 before 2019 according to panel (a) in Figure 1, the difference in difference corresponds to about 70% of the pre-treatment difference.

Coefficient estimate in column (2) of Table A.3 reveals that Twitter science stars received an additional 1.06 annual citations per publication following the publication of the top 100 Twitter scientists list by Travis (2014), hence an effect that is similar in size as that of the COVID-19. It is important to note that our difference in differences analyses consider 2011-2015 and 2016-2022 for Science's list and COVID-19 effects, respectively, so that the time coverage of the two analyses does not overlap. As a result, the Science's list effect can be considered to be a rather short-term effect so that they do not affect the pre-treatment parallel trends in the analysis for the COVID-19's effect on annual citations.

When treatment is the K-index, we find statistically significant difference-in-differences for the COVID-19 period. However, the results cannot be interpreted in a causal manner. Concerning the COVID-19 treatment, the data violates the parallel trends assumption. Prior to the treatment, citations of Twitter science stars with a high K-index were on a negative trend compared to their co-authors. Concerning the treatment of the publication of the Science list in 2014, the effects on the K-index are insignificant. We conclude that being a Twitter science star pays off while being a Kardashian does not.

4 Conclusion

Comparing Twitter scientists' annual incoming citations to those of their coauthors in COVID-related research after January 2020, we find significant difference between annual citations that are received by Twitter stars and their coauthors' papers. We find that Twitter science star status added about 1.07 extra citations following the breakout of the COVID-19 per year per article to the already existing citation gap between the Twitter scientist and their coauthors. This addition corresponds to about 70% of the pre-treatment difference. The increase in difference was 1.06 when Science's Twitter science stars list was published in 2014. Our finding shows a causal link between high visibility in social media and how this translates to actual academic merit on its own. We also find that Kardashians, i.e. scientists with an unjustified high social media attention measured by the ratio of Twitter followers to citations, do no gain from additional social media attention caused by the pandemic or the publication of the Science list.

Social media is a showcase not only for individual research papers but also for researchers, in such a way that a researcher can actively affect his or her perception by the public as an expert by sustaining an image of being scientifically more proficient than peers. It is important to note that Twitter science stars get consistently more citations than their peers which implies their celebrity status is rather well-earned. Thus one can talk about a Matthew effect in the sense that those experts who enjoy higher visibility in social media are further rewarded with more citations because of their visibility although their coauthors may have as much, if not more, expertise in the field. Researchers, research institutes, and research admins need to take their social media visibility very seriously as invisibility in social media can hurt an institution or a researcher in real terms eventually.

References

Chan, H. F., A. S. Önder, S. Schweitzer, and B. Torgler (2023): "Twitter and citations," *Economics Letters*, 231, 111270.

Hall, N. (2014): "The Kardashian index: a measure of discrepant social media profile for scientists," *Genome Biology*, 15, 1–3.

Lavezzolo, S., L. Ramiro, and P. Fernandez-Vazquez (2022): "Technocratic attitidues in COVID-19 times: Change and preference over types of experts," *European Journal of Political Research*, 61, 1123–1142.

Travis, J. (2014): "Twitter's science stars, the sequel," *Science*, 6.

You, J. (2014): "Who are the science stars of Twitter?" *Science*, 345, 1440–1441.

Appendices

A Supporting Material

Table A.1: List of Twitter science stars and their coauthors with individual K-index in parantheses

Twitter science star	Coauthor
Ben Goldacre (646)	Lahiru Handunnetthi (0.1)
Brian Krueger (21)	Rolf Renne (0)
Daniel MacArthur (29)	David V. Erbe (0)
	Monkol Lek (2)
David Eagleman (102)	Ramiro Salas (0.1)
	Vani Pariyadath (0)
Eric Topol (352)	Nicholas J. Schork (0)
	Rachel E Meyers (0)
J. Craig Venter (32)	Amalio Telenti (0.8)
	Jonathan H. Badger (0.03)
Jonathan Eisen (46)	David A. Coil (4.3)
Matt Lieberman (28)	Elizabeth Crabb Breen (0)
	Naomi I. Eisenberger (0)
Michael Eisen (40)	Jacqueline E. Villalta (0)
	Xiaoyong Li (0)
Pascal Wallisch (59)	Frédéric Chavane (0)
Robert Winston (51)	Nicholas John Dibb (0)
Simon Baron-Cohen (32)	Barbara Jacquelyn Sahakian (0)
	Bhismadev Chakrabarti (4)
Vaughan Bell (106)	n/a
Ves Dimov (43)	Frank J. Eidelman (0)

Table A.2: Differences in annual citations when COVID-19 arrived, based on publications between 2002 and 2015

	Treatment: Twitter science star			Treatment: Kardashian Index		
	(1)	(2)	(3)	(4)	(5)	(6)
Post2019XTreatment	0.0657^{c}	0.0711^{b}		0.0166^{b}	0.0168^{b}	
	[0.0337]	[0.0357]		[0.00805]	[0.00851]	
2016 × c.treated			-0.0515			-0.0182^{c}
			[0.0416]			[0.00972]
2017 × c.treated			-0.0498			-0.0129
			[0.0394]			[0.0101]
2018 × c.treated			(omitted)			(omitted)
2019 × c.treated			0.0329			0.000695
			[0.0345]			[0.00845]
2020 × c.treated			0.0694^{c}			0.0109
			[0.0384]			[0.0101]
2021 × c.treated			0.0118			-0.000728
			[0.0358]			[0.00874]
2022 × c.treated			0.0823^{c}			0.0176
			[0.0467]			[0.0114]
Indiv.Controls	Yes	Yes	Yes	Yes	Yes	Yes
Indiv.FE	No	Yes	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes
Pub.Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6615	6615	6615	6615	6615	6615
R^2	0.438	0.481	0.481	0.435	0.481	0.481

Standard errors in brackets. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$ Dependent variable is the logarithm of annual citation inflow for each scientist's publication. Columns (3) and (6) provide check for parallel trends assumption when treatment is being a top 100 Twitter star and the value of K-Index, respectively.

Table A.3: Differences in annual citations when Science published the top 100 list of Twitter science stars in 2014, based on publications between 2002 and 2011

	Treatment: Twitter science star			Treatment: Kardashian Index		
	(1)	(2)	(3)	(4)	(5)	(6)
Post2014XTreatment	0.0642^c	0.0607^c		0.0192	0.0188	
	[0.0380]	[0.0342]		[0.0127]	[0.0114]	
2011 × c.treated			-0.0956			-0.0319
			[0.0660]			[0.0227]
2012 × c.treated			-0.0346			-0.0156
			[0.0467]			[0.0158]
2013 × c.treated			(omitted)			(omitted)
2014 × c.treated			0.0292			0.00637
			[0.0440]			[0.0149]
2015 × c.treated			0.0438			0.0128
			[0.0635]			[0.0214]
2016 × c.treated			-0.0175			-0.00847
			[0.0611]			[0.0201]
Indiv.Controls	Yes	Yes	Yes	Yes	Yes	Yes
Indiv.FE	No	Yes	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	No	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes
Pub.Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3828	3828	3828	3828	3828	3828
R^2	0.513	0.536	0.536	0.513	0.536	0.536

Standard errors in brackets. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$ Dependent variable is the logarithm of annual citation inflow for each scientist's publication. Columns (3) and (6) provide check for parallel trends assumption when treatment is being a top 100 Twitter star and the value of K-Index in 2014, respectively.

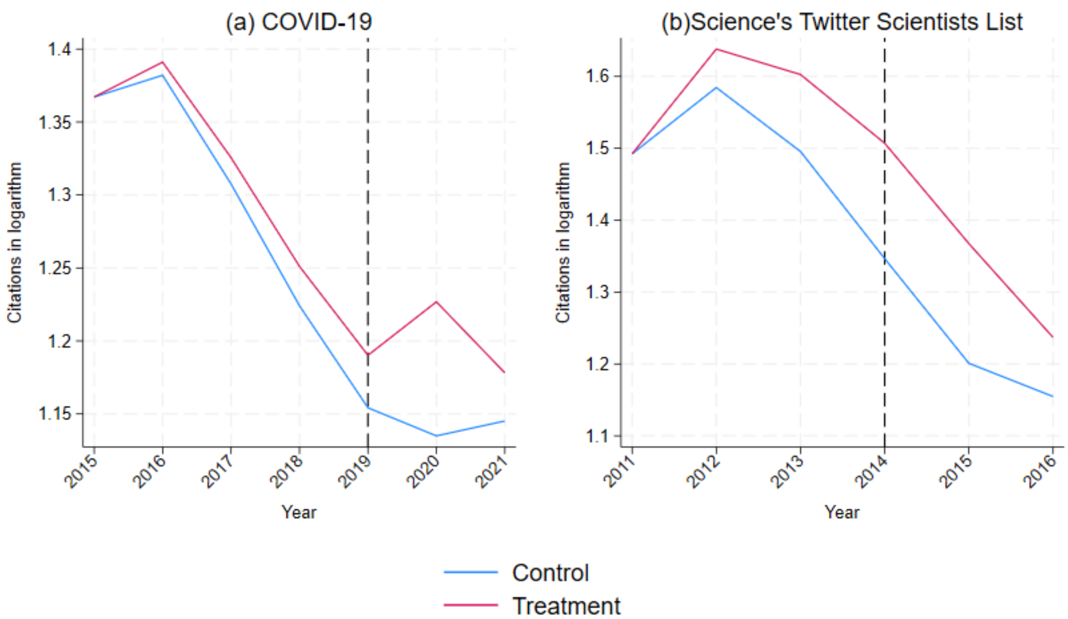

Figure A.1: Linear trends models to test parallel trends for DiD analysis at 2019 (COVID-19 in panel a) and at 2014 (Science's Twitter Scientists List in panel b).