




# Measuring the influence of non-scientific features on citations

Stefano Mammola<sup>1,2</sup>  · Elena Piano<sup>3</sup> · Alberto Doretto<sup>4</sup> · Enrico Caprio<sup>3</sup> · Dan Chamberlain<sup>3</sup>

Received: 29 December 2021 / Accepted: 20 May 2022 / Published online: 7 June 2022  
© The Author(s) 2022

## Abstract

Citations remain a prime, yet controversial, measure of academic performance. Ideally, how often a paper is cited should solely depend on the quality of the science reported therein. However, non-scientific factors, including structural elements (e.g., length of abstract, number of references) or attributes of authors (e.g., prestige and gender), may all influence citation outcomes. Knowing the predicted effect of these features on citations might make it possible to ‘game the system’ of citation counts when writing a paper. We conducted a meta-analysis to build a quantitative understanding of the effect of similar non-scientific features on the impact of scientific articles in terms of citations. We showed that article length, number of authors, author experience and their collaboration network, Impact Factors, availability as open access, online sharing, different referencing practice, and number of figures all exerted a positive influence on citations. These patterns were consistent across most disciplines. We also documented temporal trends towards a recent increase in the effect of journal Impact Factor and number of authors on citations. We suggest that our approach can be used as a benchmark to monitor the influence of these effects over time, minimising the influence of non-scientific features as a means to game the system of citation counts, and thus enhancing their usefulness as a measure of scientific quality.

**Keywords** Article impact · Scientometrics · Bibliometrics · Meta-analysis

---

✉ Stefano Mammola  
stefano.mammola@helsinki.fi; stefano.mammola@cnr.it

<sup>1</sup> LIBRe—Laboratory for Integrative Biodiversity Research, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

<sup>2</sup> Molecular Ecology Group, Water Research Institute (IRSA), National Research Council of Italy (CNR), Verbania Pallanza, Italy

<sup>3</sup> Department of Life Science and System Biology, University of Turin, Turin, Italy

<sup>4</sup> Department of Science and Technological Innovation, University of Piemonte Orientale, Alessandria, Italy

## Introduction





Most scientific careers are forged through the accumulation of papers and citations (Freeling et al., 2019). Not without controversy (Aksnes et al., 2019; Davies et al., 2021; McNutt, 2014; Todd & Ladle, 2008), metrics such as the number of publications and the number of citations remain the agreed-upon measures of scientific success. For example, in our country (Italy), the academic habilitation for Associate and Full Professor is essentially based on three criteria: number of citations, H-Index, and total number of publications. This evaluation of careers by numbers generates pressure upon academics to constantly publish high-impact papers (Smaldino & McElreath, 2016). As a result, the ability to publish and attract citations is becoming an essential academic skill.

Ideally, the impact of a scientific publication should depend solely on the quality of the science reported therein. In scientometric studies, scientific quality is typically captured by analysing factors such as the importance of the topic, level of evidence, novelty, study design, and methodology (Tahamtan & Bornmann, 2019; Tahamtan et al., 2016; Xie et al., 2022). However, evidence is accumulating that several non-scientific features contribute to the short- to long-term reach and impact of any given paper, including stylistic choices (Heard et al., 2022; Letchford et al., 2016; Martínez & Mammola, 2021; Murphy et al., 2019), number of authors (Fox et al., 2016), biases in authors' language and gender (Andersen et al., 2019; Fu & Hughey, 2019), and availability as a preprint (Fu & Hughey, 2019). By capitalizing on the conclusions of similar analyses, researchers may be tempted to create their own checklist of *dos and don'ts* to maximise publication impact.

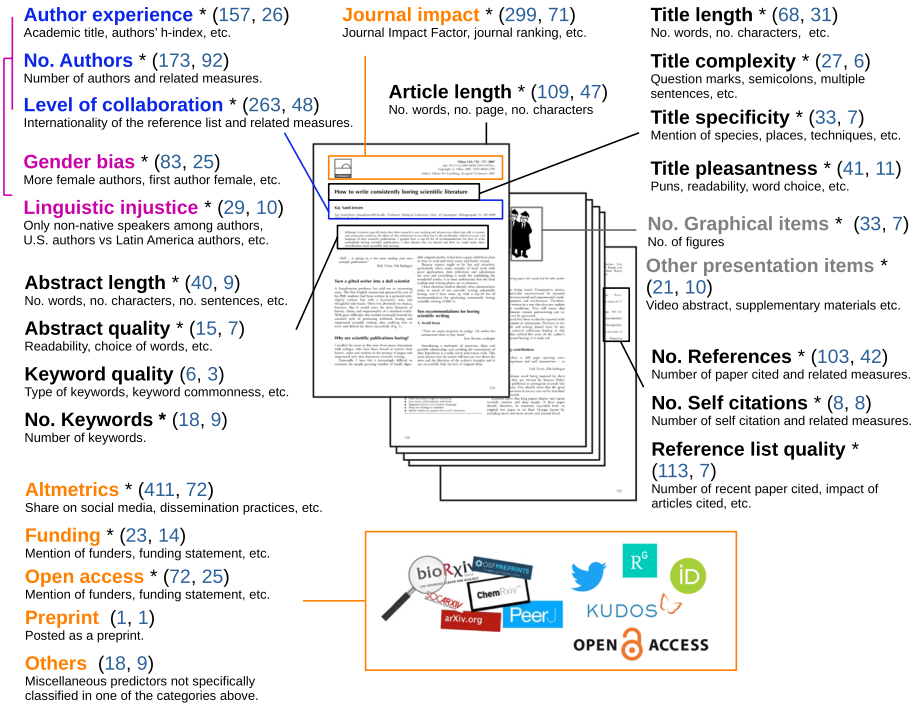
In recent years, there have been a number of seminal literature reviews that have attempted to organise this constantly growing body of literature in a digestible way (Bornmann & Daniel, 2008; Tahamtan & Bornmann, 2019; Tahamtan et al., 2016). Furthermore, there are existing meta-analyses that have looked at the influence of single features on citations, such as paper length (Xie et al., 2019) or level of collaboration (Shen et al., 2021). However, the picture is still rather crude, insofar as the effects of these non-scientific features are often weak, and their directions are likely to vary across studies, disciplines, and datasets. Two important questions naturally arise. First, can we reach a quantitative consensus on what are the most important non-scientific features that correlate with citation counts? Second, and perhaps most importantly, to what extent should we strive to achieve scientific impact by focusing our attention on these non-scientific shortcuts?

We examined these two complementary questions by undertaking a meta-analysis across the scientometric literature, designed to build a quantitative understanding of the effect of non-scientific features on citations of scientific articles. We focus on citation count as this is amongst the commonest metrics by which scientific outputs are measured, and it is also the most common *proxy* analysed in the scientometric literature (Fig. 1a). After a systematic literature search in the Web of Science (Supplementary material Table S1), we identified 262 publications testing for the importance of non-scientific features on the impact of scientific papers. We assigned each non-scientific feature to one of twenty-five major response categories across five broader general groups: *Writing features and style*, *Graphical elements*, *Pre/post-publication practices*, *Authority*, and *Authority bias* (Fig. 1b). We first looked at the consensus across our database on the impact of non-scientific features on the number of citations. As the entire dataset covered a wide range of disciplines and temporal spans, we also tested whether there has been temporal variation in the different effects and whether these effects varied across disciplines. If citations are to be used as a measure of scientific quality, then the influence of non-scientific article features on

**(a) Response – Measures of article quality/success**

<p><b>Citations *</b> (2437, 248)</p>  <p>Any measure of citations over time (e.g., no. of citations, no. of citation normalized by year, No. of early citations).</p>	<p><b>Altmetrics</b> (194, 24)</p>  <p>Any measure of attention beyond citations (e.g., reviews on Faculty of 1000, mentions on twitter, no. downloads, Altmetric score).</p>	<p><b>Impact Factor</b> (39, 15)</p>  <p>Any measure of impact of the journal in which a given paper is published (e.g., Impact factor, Ranking, prestige).</p>	<p><b>Peer-review success</b> (34, 6)</p>  <p>Any measure at the peer-review level (e.g., probability of acceptance, review scores, peer-review outcome, etc.)</p>
---	--	--	---

**(b) Predictors – Non-scientific features affecting Response**



**Author experience \*** (157, 26)  
Academic title, authors' h-index, etc.

**Journal impact \*** (299, 71)  
Journal Impact Factor, journal ranking, etc.

**Title length \*** (68, 31)  
No. words, no. characters, etc.

**No. Authors \*** (173, 92)  
Number of authors and related measures.

**Article length \*** (109, 47)  
No. words, no. page, no. characters

**Title complexity \*** (27, 6)  
Question marks, semicolons, multiple sentences, etc.

**Level of collaboration \*** (263, 48)  
Internationality of the reference list and related measures.

**Gender bias \*** (83, 25)  
More female authors, first author female, etc.

**Linguistic injustice \*** (29, 10)  
Only non-native speakers among authors, U.S. authors vs Latin America authors, etc.

**Title specificity \*** (33, 7)  
Mention of species, places, techniques, etc.

**Abstract length \*** (40, 9)  
No. words, no. characters, no. sentences, etc.

**Title pleasantness \*** (41, 11)  
Puns, readability, word choice, etc.

**Abstract quality \*** (15, 7)  
Readability, choice of words, etc.

**No. Graphical items \*** (33, 7)  
No. of figures

**Keyword quality** (6, 3)  
Type of keywords, keyword commonness, etc.

**Other presentation items \*** (21, 10)  
Video abstract, supplementary materials etc.

**No. Keywords \*** (18, 9)  
Number of keywords.

**No. References \*** (103, 42)  
Number of paper cited and related measures.

**No. Self citations \*** (8, 8)  
Number of self citation and related measures.

**Reference list quality \*** (113, 7)  
Number of recent paper cited, impact of articles cited, etc.


**Altmetrics \*** (411, 72)  
Share on social media, dissemination practices, etc.

**Funding \*** (23, 14)  
Mention of funders, funding statement, etc.

**Open access \*** (72, 25)  
Mention of funders, funding statement, etc.

**Preprint** (1, 1)  
Posted as a preprint.

**Others** (18, 9)  
Miscellaneous predictors not specifically classified in one of the categories above.



**Legend**

\* Feature analysed (no. articles  $\geq 5$ ) (no. estimates, no. articles). Note that sample sizes in (b) only concern the response variable Citation.

**Authority / Authority bias**  
**Writing features and style**  
Graphical elements  
**Pre/post publication practices**

**Fig. 1** Summary of the study design and the sampled literature. **a** Response variables, namely the measures of article impact (note that due to reduced sample size, only citations were analysed). **b** Predictor variables, namely all the non-scientific features that may affect the response. Colour coding refers to the grouping of variables in five broad categories. Only those features considered in more than four articles were analysed. Sample size in (b) are relative to the dependent variable citation. Note that the variable “Others” was not analysed despite being featured in more than 5 articles. (Color figure online)

citation counts needs to be minimised. A further objective of this study was thus to develop a method by which such influences can be assessed and monitored.

## Methods

### Literature search

We conducted a systematic review of science of science literature investigating the effect of non-scientific features on article impacts using the Web of Science Core Collection database over all citation indices, all document types, all years, and all languages (queries were made between 01 and 03 July 2020). An overview of the plan for the meta-analysis is reported in Supplementary material Box S1 and a PRISMA diagram (Moher et al., 2009; Page et al., 2021) is given in Supplementary material Fig. S1. Based on background knowledge of factors that may affect an article's impact (Tahamtan et al., 2016), we conducted separate queries for different non-scientific features relating to Title, Abstract, Keywords, Main text, Figures, Author-list, Bibliography, Publication strategy, and Post-publication strategy (Supplementary material Table S1; these non-scientific features were subsequently grouped in 25 major response categories across 5 broader general groups as in Fig. 1b).

We initially screened titles and abstracts to exclude clearly inappropriate references. In addition, we applied a series of stricter exclusion criteria (Supplementary material Box S1), eliminating: (i) purely descriptive papers (e.g., Editorial and Opinion pieces); (ii) research articles not quantifying article impact numerically; (iii) studies focusing solely on variables related to the scientific content (e.g., importance of the topic, novelty, study design, methodology); and (iv) articles that concerned trends over time with no reference to article impact/quality. Tests of inter-rater agreement between authors with Cohen's kappa (Cohen, 1960) showed a good level of repeatability of study selection using the specified criteria for all searches (Supplementary material Table S1). Further details of the search procedure and the repeatability analysis are given in Supplementary text.

Following the screening phase, we extracted predictors and response variables for each paper, as well as relevant meta-data (discipline, sample size, number of journals analysed, minimum, maximum, and range of publication years considered). At this stage, we identified several studies that could not be used for the final analysis due to inadequate presentation of results. However, we contacted corresponding authors of these studies asking for missing information, which yielded further usable data from 19 studies (response rate = 58.5%).

### Effect size calculation

We used only the number of citations as our dependent variable as this is the main currency for assessing perceived quality of scientific literature (Aksnes et al., 2019). All other response variables lacked a sufficient sample size (Fig. 1a).

We converted test statistics that described the link between citations and non-scientific features of papers to Pearson's  $r$  using standard conversion formulas (Lajeunesse, 2013). Pearson's  $r$  expresses the effect size, i.e. the strength of a given linear association between the number of citations and non-scientific features. It ranges continuously between  $-1.0$  and  $1.0$ , where positive values indicate a positive effect of the feature on citations, values close to 0 no significant association, and negative values negative effects. We calculated

Pearson's  $r$  for any comparison within a study, such as when different non-scientific features or different levels for factors were analysed by the authors.

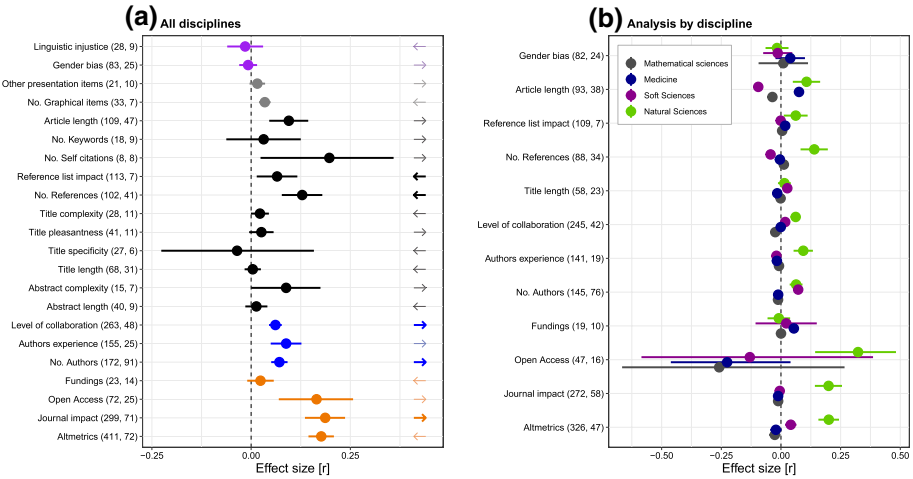
## Meta-analysis

We conducted analyses in R version 4.0.3 (R Core Team, 2021), using the 'metafor' package version 2.4.0 (Viechtbauer, 2010). To approximate normality, we converted Pearson's  $r$  to Fisher's  $z$  (Rosenberg et al., 2013). However, as in Chamberlain et al. (2020), for the presentation of the results, we back-transformed Fisher's  $z$  values and their 95% confidence intervals to Pearson's  $r$  to ease visualisation. We interpreted model-derived estimates of Pearson's  $r$  as the strength of the standardized effect, which was considered significant when the 95% confidence intervals did not overlap zero.

In the first set of meta-analytic linear mixed-effects models ('metafor' function *rma.mv*), we assessed the extent to which different non-scientific features affect the number of citations across the whole sample. In all models, we specified a publication-level nesting factor to account for study-level non-independence due to multiple measurements per study (mean  $\pm$  s.d. =  $9.4 \pm 34.1$  estimates/publication). Many journals have fixed formats (e.g., abstract length, number of keywords). Some of our results may have arisen because journals that have particular formats, especially in terms of writing features and style, also happen to be cited more. To verify the robustness of our approach, we therefore repeated the first analysis on a subset of scientometric papers that only focused on a single journal. This ensured that the estimates derived from this subset of papers were controlled for journal-specific features such as Impact Factor, ranking, field or international visibility.

Subsequently, we tested the influence of two moderators (that is, variables that condition the effect size in meta-analyses) on the citation effect. We explored temporal effects by running a set of models that included time as a moderator for each predictor variable, aiming to establish whether the direction of the effect was moving towards zero (namely, toward a lack of effect) or away from zero (increased positive or negative effects) in recent years. To this end, we included the minimum year of the database considered in the publication as a moderator in each model. This way, we could see if the effect is different between analyses based on older versus recent samples of literature. Note that we also repeated the analysis using maximum year of publication instead of minimum year of publication, and results were similar to those using the minimum year (Supplementary material Fig. S3). The direction of the effect was based on the direction of the estimate for each moderator. We considered cases where the effect of year was significant as evidence of a significant temporal effect (bold arrows in Fig. 2a). As a general indication, we also reported the direction for non-significant effects.

The second moderator we tested was discipline, given that the effect of non-scientific features on citations may vary in strength and direction across scientific domains (Tahamtan et al., 2016). As discipline is a categorical variable, we constructed hierarchical models that tested for differences between the levels of discipline within a given model (between-group heterogeneity). We considered cases where between-group heterogeneity was significant as evidence of a moderator effect. To balance the sample size across levels, we grouped disciplines into four levels: (i) *Mathematical sciences* (Engineering, Informatics, Mathematics, Physics;  $N=718$  estimates); (ii) *Medicine* ( $N=718$ ); (iii) *Natural sciences* (all biological disciplines and chemistry;  $N=539$ ); and (iv) *Soft sciences* (all humanistic disciplines, including Law, Politics, and Economy;  $N=317$ ). In all of these models,



**Fig. 2** Results of the meta-analysis. Estimates of the effect size of non-scientific features on article citations, expressed as standardized Pearson’s  $r \pm 95\%$  confidence limits. Sample sizes are given in parentheses (number of standardized estimates, number of studies). **a** The overall effect is the strength of the effect without moderators. Arrows on the right indicate the direction of the temporal effect, based on a second model which included the minimum of the publications considered in a given study as a moderator. Bold arrows denote significant temporal trends ( $p < .001$ ). Colour coding as in Fig. 1b. Model estimates and exact p-values are given in Table 1. **b** The overall effect is the strength of the effect including discipline as a moderator. Model estimates are given in Table 2

we excluded estimates deriving from articles that did not make a distinction across disciplines ( $N = 204$ ).

**Publication bias**

Publication bias arises when studies finding negative results are less likely to be published than those finding supportive evidence—a phenomenon also termed the “file-drawer effect” because negative results are imagined to rest, unread, in scientists’ drawers. We evaluated publication bias via fail-safe number analysis, as implemented in the ‘metafor’ function *fsn*. Specifically, we used Rosenthal’s method (Rosenberg, 2005; Rosenthal, 1979) to calculate the number of studies averaging negative results that would have to be added to the given set of observed outcomes (predictors) to reduce the combined significance level to a target alpha level of 0.05. According to the fail-safe N analysis (Table 1), there was no evidence of publication bias in any model except for the Number of keywords and Title pleasantness, which therefore must be interpreted with caution.

**Results and discussion**

**Influence of non-scientific features on citations**

We identified 2,312 studies in the initial literature search (Supplementary material—Appendix S2), of which 997 were deemed relevant for testing for the effect of non-scientific

**Table 1** Estimated model parameters (full model and model with year as moderator)

Predictor	Full model				Partial model with year [estimates for the moderator]				Validation [Rosenthal (1979) approach]	
	N	Beta±SE	CI	p val	Beta±SE	CI	p val	Fail-safe number	p val	
Altmetrics	411	0.36±0.03	(0.29, 0.42)		0.0012±0.00039	(- 0.00545, 0.00785)	0.72353	32,317,091	<0.001	
Journal impact	299	0.38±0.05	(0.27, 0.48)		- 0.00148±0.00001	(- 0.00247, 0.00049)	0.00337	3,900,888	<0.001	
Open access	72	0.33±0.1	(0.14, 0.53)		- 0.00899±0.01888	(- 0.046, 0.02802)	0.6341	266,714	<0.001	
Fundings	23	0.05±0.03	(- 0.02, 0.11)		0.00167±0.00202	(- 0.00229, 0.00562)	0.40936	2952	<0.001	
No. authors	172	0.14±0.02	(0.10, 0.18)		- 0.00154±0.00038	(- 0.00229, 0.00001)	<0.001	2,920,500	<0.001	
Authors experience	155	0.18±0.04	(0.10, 0.25)		- 0.0048±0.00298	(- 0.01064, 0.00103)	0.10683	6,581,752	<0.001	
Level of collaboration	263	0.12±0.02	(0.09, 0.15)		- 0.01374±0.00046	(- 0.01464, 0.01285)	<0.001	4,951,430	<0.001	
Abstract length	40	0.03±0.03	(- 0.03, 0.08)		0.00344±0.00686	(- 0.01, 0.01689)	0.61567	12,777	<0.001	
Abstract complexity	15	0.18±0.09	(0.00, 0.35)		- 0.00655±0.00604	(- 0.01839, 0.00529)	0.27827	11,921	<0.001	
Title length	68	0.01±0.02	(- 0.03, 0.05)		0.00073±0.00156	(- 0.00234- 0.0038)	0.64071	38,122	<0.001	
Title specificity	27	- 0.07±0.2	(- 0.46, 0.32)		0.00781±0.01457	(- 0.02075, 0.03637)	0.59184	303	<0.001	
Title pleasantness	41	0.05±0.03	(- 0.01, 0.11)		- 0.00016±0.00249	(- 0.00504, 0.00472)	0.94834	0	0.173	
Title complexity	28	0.04±0.02	(0.00, 0.09)		0.00032±0.00178	(- 0.00316, 0.00338)	0.85745	381	<0.001	
No. References	102	0.26±0.05	(0.15, 0.36)		0.0079±0.00068	(0.00657, 0.00923)	<0.001	4,189,775	<0.001	
Reference list impact	113	0.13±0.05	(0.03, 0.23)		0.00508±0.00047	(0.00416, 0.00599)	<0.001	570,999	<0.001	
No. self citations	8	0.4±0.18	(0.05, 0.75)		- 0.01815±0.01258	(- 0.04281, 0.0065)	0.14904	480,112	<0.001	
No. keywords	18	0.06±0.1	(- 0.13, 0.25)		- 0.0113±0.01228	(- 0.03537, 0.01277)	0.35741	0	0.274	
Article length	109	0.19±0.05	(0.09, 0.29)		- 0.00046±0.00396	(- 0.00822, 0.0073)	0.90676	683,072	<0.001	
No. graphical items	33	0.07±0.01	(0.04, 0.10)		0.00159±0.00165	(- 0.00164, 0.00482)	0.33445	1557	<0.001	
Other presentation items	21	0.03±0.02	(- 0.01, 0.07)		- 0.00273±0.0026	(- 0.00783, 0.00237)	0.2945	189	<0.001	
Gender bias	83	- 0.02±0.02	(- 0.06, 0.03)		- 0.00158±0.00134	(- 0.00421, 0.00106)	0.2402	19,680	<0.001	
Linguistic injustice	28	- 0.03±0.05	(- 0.12, 0.06)		0.0034±0.00428	(- 0.00499, 0.01178)	0.42697	625	<0.001	

Estimated model parameters for the full model, for the partial model with year as a moderator, and model validation using the Rosenthal approach. Colour coding as in Fig. 1b



features on article impact. A subset of 262 studies satisfied the necessary inclusion criteria for the meta-analysis (Supplementary material—Fig. S1). Number of citations was the sole response variable with sufficient sample size, namely 250 articles and 2,361 unique standardized estimates (Fig. 1a).

Overall, effects of non-scientific features on article citations were consistently weak—all standardized Pearson's  $r < \pm 0.2$  (Fig. 2a). This is in line with the general understanding that variables related to the scientific content (e.g., importance of the topic, level of evidence, novelty, study design, methodology) rather than non-scientific features, are the most important factors capturing variance in citation outcomes (Tahamtan et al., 2016). However, several effects were highly significant, suggesting that many non-scientific features can contribute to inflating an article's citation rate.

There is an implicit assumption that the variables we have analysed are not genuine measures of scientific quality, but it could be argued that a few are. In other words, some of the factors considered here may sometimes be plausible surrogates of 'quality' (e.g., author experience, collaboration measures, Impact Factor), whereas most others are just artefacts of the publishing and citation system (e.g., reference lists, article length, number of keywords, self-citations).

Among the non-scientific features that might be at least partially related to scientific quality, those referring to journal impact and author experience had the strongest positive effects. Publishing in prestigious journals with higher Impact Factors gives a citation advantage, possibly due to greater visibility and perceived trust in the science appearing in top-tier journals (Callahan et al., 2002). Furthermore, papers with more authors, with more experienced authors and with authors having broader collaborative networks tended to be more highly cited. This finding matches the results of an independent meta-analysis (Shen et al., 2021), and may be explained by multiple factors. On the one hand, the quality of research is likely to be higher when multi-disciplinary teams and experienced authors are working together (Cardoso et al., 2021; Falkenberg & Tubb, 2017; James Jacob, 2015). On the other hand, authors that are famous, highly-cited and/or more advanced in their career may achieve more citations due to their prestige and perceived credibility in the field (Tahamtan et al., 2016). Furthermore, higher citation counts may simply be the result of a greater visibility arising from a multitude of co-authors and their respective networks (Bosquet & Combes, 2013).

Further positive effects were associated with different features of reference lists, including the reference list length, the overall impact of the literature cited in a given paper and the number of self-citations. The positive effect of self-citations on the number of citations suggests that excessive self-citations may be used by some authors as a way to increase their visibility and boost their citation metrics (Fowler & Aksnes, 2007). However, self-citations are also an integral part of scientific progress, reflecting the cumulative nature of individual research (Glänzel & Thijs, 2004; Ioannidis, 2015; Mishra et al., 2018; Penders, 2018). Papers with a high number of self-citations that are part of a long-term research line are often more visible and citable, and this may lead to accumulating more citations (Mammola et al., 2021) (Table 2).

Additional positive significant effects refer to factors that are likely to be mostly structural, including article length and number of graphical items. These results match those of previous studies (Tahamtan & Bornmann, 2019; Tahamtan et al., 2016; Xie et al., 2019). This effect may arise because longer papers that cite more articles and that include more figures may be addressing a greater diversity of ideas and topics (Ball, 2008; Elgendi, 2019; Fox et al., 2016). Indeed, there is evidence that, in recent years, individual papers are becoming more densely packed with information and thus may contain more citable



**Table 2** Estimated model parameters (model with discipline as moderator)

Predictor	Names	<i>N</i>	Beta ± SE	CI	<i>p</i> val
Altmetrics	Intercept (natural sciences)	326	0.41 ± 0.05	(0.32, 0.5)	< 0.001
	Soft sciences		0.08 ± 0.02	(0.03, 0.13)	0.00079
	Mathematical sciences		− 0.05 ± 0.02	(− 0.1, 0)	0.03347
	Medicine		− 0.04 ± 0.02	(− 0.09, 0.01)	0.08032
Journal impact	Intercept (natural sciences)	272	0.4 ± 0.06	(0.29, 0.52)	< 0.001
	Soft sciences		− 0.01 ± 0.01	(− 0.02, 0)	0.04172
	Mathematical sciences		− 0.02 ± 0	(− 0.03, − 0.02)	< 0.001
	Medicine		− 0.02 ± 0	(− 0.03, − 0.02)	< 0.001
Open access	Intercept (natural sciences)	47	0.67 ± 0.19	(0.29, 1.05)	0.00058
	Soft sciences		− 0.26 ± 0.55	(− 0.34, 0.81)	0.63267
	Mathematical sciences		− 0.53 ± 0.55	(− 0.61, 0.55)	0.33445
	Medicine		− 0.46 ± 0.28	(− 0, 0.08)	0.09526
Fundings	Intercept (Natural sciences)	19	− 0.02 ± 0.05	(− 0.11, 0.08)	0.71034
	Soft sciences		0.04 ± 0.13	(− 0.21, 0.3)	0.7365
	Mathematical sciences		0 ± 0.01	(− 0.02, 0.02)	0.96718
	Medicine		0.11 ± 0.02	(0.08, 0.14)	< 0.001
No. authors	Intercept (natural sciences)	146	0.13 ± 0.03	(0.07, 0.18)	< 0.001
	Soft sciences		0.14 ± 0	(0.14, 0.15)	< 0.001
	Mathematical sciences		− 0.02 ± 0	(− 0.03, − 0.02)	< 0.001
	Medicine		− 0.02 ± 0	(− 0.03, − 0.02)	< 0.001
Authors experience	Intercept (natural sciences)	143	0.18 ± 0.04	(0.1, 0.26)	< 0.001
	Soft sciences		− 0.04 ± 0.01	(− 0.05, − 0.03)	< 0.001
	Mathematical sciences		− 0.02 ± 0	(− 0.02, − 0.01)	< 0.001
	Medicine		− 0.03 ± 0	(− 0.04, − 0.03)	< 0.001
Level of collaboration	Intercept (natural sciences)	245	0.12 ± 0.02	(0.09, 0.16)	< 0.001
	Soft sciences		0.04 ± 0	(0.03, 0.04)	< 0.001
	Mathematical sciences		− 0.05 ± 0	(− 0.05, − 0.04)	< 0.001
	Medicine		0 ± 0	(0, 0)	0.23472
Title length	Intercept (natural sciences)	58	0.03 ± 0.03	(− 0.03, 0.08)	0.29512
	Soft sciences		0.05 ± 0	(0.05, 0.06)	< 0.001
	Mathematical sciences		0 ± 0	(− 0.01, 0)	0.4
	Medicine		− 0.03 ± 0	(− 0.04, − 0.02)	< 0.001
No. references	Intercept (natural sciences)	89	0.27 ± 0.06	(0.16, 0.39)	< 0.001
	Soft sciences		− 0.09 ± 0	(− 0.09, − 0.08)	< 0.001
	Mathematical sciences		0.02 ± 0	(0.02, 0.03)	< 0.001
	Medicine		− 0.01 ± 0	(− 0.02, 0)	0.00497
Reference list impact	Intercept (natural sciences)	109	0.12 ± 0.05	(0.02, 0.22)	0.01548
	Soft sciences		0 ± 0	(− 0.01, 0)	0.39807
	Mathematical sciences		0.01 ± 0	(0.01, 0.01)	< 0.001
	Medicine		0.04 ± 0	(0.03, 0.04)	< 0.001
Article length	Intercept (natural sciences)	93	0.22 ± 0.06	(0.1, 0.33)	0.00027
	Soft sciences		− 0.19 ± 0	(− 0.2, − 0.18)	< 0.001
	Mathematical sciences		− 0.07 ± 0	(− 0.08, − 0.07)	< 0.001
	Medicine		0.15 ± 0	(0.15, 0.16)	< 0.001

**Table 2** (continued)

Predictor	Names	<i>N</i>	Beta ± SE	CI	<i>p</i> val
Gender bias	Intercept (natural sciences)	82	$-0.03 \pm 0.05$	(- 0.13, 0.06)	0.50084
	Soft sciences		$-0.03 \pm 0.06$	(- 0.15, 0.1)	0.66385
	Mathematical sciences		$0.02 \pm 0.11$	(- 0.19, 0.23)	0.85204
	Medicine		$0.08 \pm 0.06$	(- 0.05, 0.2)	0.21739

Colour coding as in Fig. 1b

information (Cordero et al., 2016). Furthermore, longer reference lists may make papers more visible in online searches, while also attracting *tit-for-tat* citations, that is, the tendency of cited authors to cite the papers that cited them (Mammola et al., 2021).

No other analysed predictors exerted significant control on citations (Fig. 2a). Among others, the lack of effect of gender may come as a surprise, given that the discourse on gender biases is timely (Abramo et al., 2021; AlShebli et al., 2020; Casad et al., 2021; Davies et al., 2021; Holman & Morandin, 2019; Kwon, 2022). Note that the effect was in the expected direction (negative), but only weakly significant as the confident interval overlapped zero (Fig. 2a). While the existence of gender inequality within Academia is undeniable, this bias may not be best captured by citations—but see Dworkin et al. (2020) for larger effects in Neuroscience. For example, a recent analysis focused on more than 1 million medicine papers published between 2008 and 2014 showed that differences in citation distributions between males and females are very small and mostly attributable to journal prestige and self-citations (Andersen et al., 2019). Indeed, once a paper is published, it is unlikely that it will be less cited based on gender considerations, in part because only the surname of authors is available in most reference managers. Conversely, gender bias in publication outcomes may be best captured by other response variables, especially those related to peer-review success (Fox & Paine, 2019) or Impact Factor (Barrios et al., 2013; Holman & Morandin, 2019), which we could not analyse due to low sample size (Fig. 1a).

Repeating the analysis for studies that considered only a single journal revealed largely consistent patterns with the main analysis (cf. Fig. 2 and Supplementary material Fig. S2), thus indicating that the observed effects operated within journals as well as between them, i.e. the observed effects were unlikely to have been due to particular formats being associated with journals that had higher citations.

## Temporal trends

There was a significant temporal effect for journal Impact Factor, whereby the positive effect of journal impact on citations is becoming stronger over time (Fig. 2a). There was also a significant temporal effect of the number of authors and level of collaboration, with more authors and broader collaborative networks exerting a stronger influence on citations in recent years. Science is indeed becoming more and more collaborative and trans-disciplinary (Knapp et al., 2015; Sahneh et al., 2021), a trend made possible by gigantic technological advances in communication enabling effective collaboration among multiple authors and large consortia—e.g., in medicine (International Human Genome Sequencing consortium, 2004) and physics (Castelvecchi, 2015).

Other significant temporal effects pertained to the decreasing influence of reference list features on citations in recent years. The fact that reference list features are less closely

associated with citations may imply better referencing practice by authors in recent years, a behaviour possibly facilitated by increasing awareness of the importance of responsible referencing (Kwon, 2022; Penders, 2018), but also to a greater availability of performing technologies to browse literature through the Internet (Gingras et al., 2009; Mammola et al., 2021).

### Discipline-specific effects

The above patterns were consistent across most disciplines, with a few exceptions (Fig. 2b, Table 2). Natural sciences largely deviated with respect to open access, journal Impact Factor and online sharing. The availability of funding positively influenced citations in medicine, a field that notoriously receives high financial support (Murphy & Topel, 2010). Furthermore, longer papers were associated with more citations in medicine and natural science, whereas more concise papers had a citation advantage in mathematical sciences and soft sciences.

### A recipe for success?

Our analysis indicates that patterns exist in the data regarding the effect of non-scientific factors on how often a particular paper is cited. Insofar as citation counts are widely used for measuring academic performance, one might be tempted to use this knowledge to game the system of citation counts when writing a paper. However, this is likely to be problematic. If citations are determined by a number of technical factors that are unrelated to scientific quality, evaluating the impact of a paper solely based on this blunt metric entails a substantial intrinsic bias—scientific quality is necessarily a multifaceted concept (Polanyi et al., 1962). This is why a modern discourse on this subject focuses on exploring new ways to express scientific quality, for example by decomposing it into fundamental components such as Solidity and Plausibility, Originality and Novelty, and Scientific Value (Aksnes et al., 2019).

An additional cause of concern is the existence of temporal trends towards an increase of some of these effects over time. This ‘rise and fall’ of some non-scientific features implies that the number of citations is influenced by factors that are deemed as important by the scientific community in some periods, but not in others. Ideally, we would want the effects in Fig. 2a to move towards the middle (a zero effect), so that factors unrelated to scientific quality can no longer be used to boost citations. However, we found that about half of the factors analysed here are shifting away from zero (that is, they are becoming more influential over time), further strengthening the idea that the number of citations alone is not enough to evaluate the scientific quality of a paper.

As a final corollary note, it is important to mention that publishing behaviours of authors are often modulated by journals and institutions. For example, most journal guidelines limit the abstract and article length, number of keywords, number of references, and rarely even the number of authors (e.g., in *Trends in Ecology and Evolution*). Furthermore, assessment agencies and institutions are increasingly using Altmetrics and citations as measures of impact. All these constraints feedback to affect the behaviour of authors, and vice versa.

## Conclusion

In conclusion, let us ask it one final time “... to achieve scientific impact, should we ‘game the system’ and exploit these non-scientific features to our advantage?” Although our meta-analysis pointed out that some non-scientific features can be used to increase the probability of being cited, the temporal variability of their relevance jeopardizes any “recipes for success” based on their use. Consequently, if we want citations to be used as a measure of scientific quality, the answer is “no”, but the fact that citations remain among the most important metric for academics, especially early career researchers, is of sufficient magnitude to warrant constant reflection on our publication practices. To this end, our method can act as a benchmark to monitor the influence of these features over time. Ideally, to adopt citation counts as a measure of scientific quality, effect sizes in Fig. 2 should be zero for most of themes addressed. This can be achieved by monitoring new literature being published on the subject to see how these effects vary in the years ahead.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11192-022-04421-7>.

**Acknowledgements** S.M. acknowledge support from the European Commission via the Marie Skłodowska-Curie Individual Fellowships program (H2020-MSCA-IF-2019; project number 882221).

**Author contributions** SM and DC conceived the idea, designed the methodology and analysed the data. All authors collected data. SM wrote the first draft and prepared figures, with substantial contributions by DC. All authors contributed to the writing with suggestions and critical comments.

**Funding** Open Access funding provided by University of Helsinki including Helsinki University Central Hospital.

**Data availability** Data and R code to reproduce the analysis are available in Figshare (doi:<https://doi.org/10.6084/m9.figshare.19426076>).

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abramo, G., Aksnes, D. W., & D’Angelo, C. A. (2021). Gender differences in research performance within and between countries: Italy vs Norway. *Journal of Informetrics*, 15(2), 101144. <https://doi.org/10.1016/j.joi.2021.101144>
- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*. <https://doi.org/10.1177/2158244019829575>

- AlShebli, B., Makovi, K., & Rahwan, T. (2020). Retraction Note: The association between early career informal mentorship in academic collaborations and junior author performance. *Nature Communications*, 11(1), 6446. <https://doi.org/10.1038/s41467-020-20617-y>
- Andersen, J. P., Schneider, J. W., Jagsi, R., & Nielsen, M. W. (2019). Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. *eLife*, 8, e45374. <https://doi.org/10.7554/eLife.45374>
- Ball, P. (2008). A longer paper gathers more citations. *Nature*, 455(7211), 274. <https://doi.org/10.1038/455274a>
- Barrios, M., Villarroya, A., & Borrego, Á. (2013). Scientific production in psychology: A gender analysis. *Scientometrics*, 95(1), 15–23. <https://doi.org/10.1007/s11192-012-0816-4>
- Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80. <https://doi.org/10.1108/00220410810844150>
- Bosquet, C., & Combes, P.-P. (2013). Are academics who publish more also more cited? Individual determinants of publication and citation records. *Scientometrics*, 97(3), 831–857. <https://doi.org/10.1007/s11192-013-0996-6>
- Callaham, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA*, 287(21), 2847–2850. <https://doi.org/10.1001/jama.287.21.2847>
- Cardoso, P., Fukushima, C. S., & Mammola, S. (2021). Quantifying the international collaboration of researchers and research institutions. *MetaArXiv Preprints*. <https://doi.org/10.31222/osf.io/b6anf>
- Casad, B. J., Franks, J. E., Garasky, C. E., Kittleman, M. M., Roesler, A. C., Hall, D. Y., & Petzel, Z. W. (2021). Gender inequality in academia: Problems and solutions for women faculty in STEM. *Journal of Neuroscience Research*, 99(1), 13–23. <https://doi.org/10.1002/jnr.24631>
- Castelvecchi, D. (2015). Physics paper sets record with more than 5,000 authors. *Nature*. <https://doi.org/10.1038/nature.2015.17567>
- Chamberlain, D., Reynolds, C., Amar, A., Henry, D., Caprio, E., & Batáry, P. (2020). Wealth, water and wild-life: Landscape aridity intensifies the urban luxury effect. *Global Ecology and Biogeography*, 29(9), 1595–1605. <https://doi.org/10.1111/geb.13122>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cordero, R. J. B., de León-Rodríguez, C. M., Alvarado-Torres, J. K., Rodríguez, A. R., & Casadevall, A. (2016). Life science's average publishable unit (APU) Has increased over the past two decades. *PLoS ONE*, 11(6), e0156983. <https://doi.org/10.1371/journal.pone.0156983>
- Davies, S. W., Putnam, H. M., Ainsworth, T., Baum, J. K., Bove, C. B., Crosby, S. C., et al. (2021). Promoting inclusive metrics of success and impact to dismantle a discriminatory reward system in science. *PLOS Biology*, 19(6), e3001282. <https://doi.org/10.1371/journal.pbio.3001282>
- Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, 23(8), 918–926. <https://doi.org/10.1038/s41593-020-0658-y>
- Elgendi, M. (2019). Characteristics of a highly cited article: A machine learning perspective. *IEEE Access*, 7, 87977–87986. <https://doi.org/10.1109/ACCESS.2019.2925965>
- Falkenberg, L. J., & Tubb, A. (2017). Undisciplined thinking facilitates accessible writing: A response to Doubleday and Connell. *Trends in Ecology & Evolution*, 32(12), 894–895. <https://doi.org/10.1016/j.tree.2017.09.009>
- Fowler, J. H., & Aksnes, D. W. (2007). Does self-citation pay? *Scientometrics*, 72(3), 427–437. <https://doi.org/10.1007/s11192-007-1777-2>
- Fox, C. W., & Paine, C. E. T. (2019). Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. *Ecology and Evolution*, 9(6), 3599–3619. <https://doi.org/10.1002/ece3.4993>
- Fox, C. W., Paine, C. E. T., & Sauterey, B. (2016). Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*, 6(21), 7717–7726. <https://doi.org/10.1002/ece3.2505>
- Freeling, B., Doubleday, Z. A., & Connell, S. D. (2019). Opinion: How can we boost the impact of publications? Try better writing. *Proceedings of the National Academy of Sciences of the United States of America*, 116(2), 341–343. <https://doi.org/10.1073/pnas.1819937116>
- Fu, D. Y., & Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer reviewed article. *Life*, 8, e52646. <https://doi.org/10.7554/eLife.52646>
- Gingras, Y., Larivière, V., & Archambault, É. (2009). Literature citations in the Internet Era. *Science*, 323(5910), 36. <https://doi.org/10.1126/science.323.5910.36>
- Glänzel, W., & Thijs, B. (2004). The influence of author self-citations on bibliometric macro indicators. *Scientometrics*, 59(3), 281–310. <https://doi.org/10.1023/B:SCIE.0000018535.99885.e9>

- Heard, S. B., Cull, C. A., & White, E. R. (2022). If this title is funny, will you cite me? Citation impacts of humour and other features of article titles in ecology and evolution. *bioRxiv*. <https://doi.org/10.1101/2022.03.18.484880>
- Holman, L., & Morandin, C. (2019). Researchers collaborate with same-gendered colleagues more often than expected across the life sciences. *PLoS ONE*, *14*(4), e0216128. <https://doi.org/10.1371/journal.pone.0216128>
- International Human Genome Sequencing consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. <https://doi.org/10.1038/nature03001>
- Ioannidis, J. P. A. (2015). A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. *Journal of Psychosomatic Research*, *78*(1), 7–11. <https://doi.org/10.1016/j.jpsychores.2014.11.008>
- James Jacob, W. (2015). Interdisciplinary trends in higher education. *Palgrave Communications*, *1*(1), 15001. <https://doi.org/10.1057/palcomms.2015.1>
- Knapp, B., Bardenet, R., Bernabeu, M. O., Bordas, R., Bruna, M., Calderhead, B., et al. (2015). Ten simple rules for a successful cross-disciplinary collaboration. *PLOS Computational Biology*, *11*(4), e1004214. <https://doi.org/10.1371/journal.pcbi.1004214>
- Kwon, D. (2022). The rise of citational justice: How scholars are making references fairer. *Nature*, *603*, 568–571.
- Lajeunesse, M. J. (2013). Recovering missing or partial data from studies: A survey of conversions and imputations for meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of meta-analysis in ecology and evolution* (pp. 195–206). Princeton University Press.
- Letchford, A., Preis, T., & Moat, H. S. (2016). The advantage of simple paper abstracts. *Journal of Informetrics*, *10*(1), 1–8. <https://doi.org/10.1016/J.JOI.2015.11.001>
- Mammola, S., Fontaneto, D., Martínez, A., & Chichorro, F. (2021). Impact of the reference list features on the number of citations. *Scientometrics*, *126*, 785–799. <https://doi.org/10.1007/s11192-020-03759-0>
- Martínez, A., & Mammola, S. (2021). Specialized terminology reduces the number of citations to scientific papers. *Proceedings of the Royal Society b: Biological Sciences*, *288*(1948), 20202581.
- McNutt, M. (2014). The measure of research merit. *Science*, *346*(6214), 1155. <https://doi.org/10.1126/science.aaa3796>
- Mishra, S., Fegley, B. D., Diesner, J., & Torvik, V. I. (2018). Self-citation is the hallmark of productive authors, of any gender. *PLoS ONE*, *13*(9), e0195773. <https://doi.org/10.1371/journal.pone.0195773>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535. <https://doi.org/10.1136/bmj.b2535>
- Murphy, K. M., & Topel, R. H. (2010). The economic value of medical research. In *Measuring the gains from medical research* (pp. 41–73). University of Chicago Press.
- Murphy, S. M., Vidal, M. C., Hallagan, C. J., Broder, E. D., Barnes, E. E., Horna Lowell, E. S., & Wilson, J. D. (2019). Does this title bug (Hemiptera) you? How to write a title that increases your citations. *Ecological Entomology*, *44*(5), 593–600. <https://doi.org/10.1111/een.12740>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. <https://doi.org/10.1136/bmj.n71>
- Penders, B. (2018). Ten simple rules for responsible referencing. *PLOS Computational Biology*, *14*(4), e1006036. <https://doi.org/10.1371/journal.pcbi.1006036>
- Polany, M., Ziman, J., & Fuller, S. (1962). The republic of science: Its political and economic theory. *Minerva*, *38*(1), 1–32.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution; International Journal of Organic Evolution*, *59*(2), 464–468.
- Rosenberg, M. S., Rothstein, H. R., & Gurevitch, J. (2013). Effect sizes: Conventional choices and calculations. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of meta-analysis in ecology and evolution* (pp. 61–71). Princeton University Press.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Sahneh, F., Balk, M. A., Kiskey, M., Chan, C., Fox, M., Nord, B., Lyons, E., Swetnam, T., Huppenkothen, D., Sutherland, W., Walls, R. L., Quinn, D. P., Tarin, T., LeBauer, D., Ribes, D., Birnie, D. P., Lushbough, C., Carr, E., Nearing, G., ... Kobourov, S. (2021/2022). Ten simple rules to cultivate transdisciplinary collaboration in data science. *PLoS Computational Biology*, *17*(5), e1008879. <https://doi.org/10.1371/journal.pcbi.1008879>

- Shen, H., Xie, J., Li, J., & Cheng, Y. (2021). The correlation between scientific collaboration and citation count at the paper level: A meta-analysis. *Scientometrics*, *126*(4), 3443–3470. <https://doi.org/10.1007/s11192-021-03888-0>
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384. <https://doi.org/10.1098/rsos.160384>
- Tahamtan, I., & Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics*, *121*(3), 1635–1684. <https://doi.org/10.1007/s11192-019-03243-4>
- Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, *107*(3), 1195–1225. <https://doi.org/10.1007/s11192-016-1889-2>
- Todd, P. A., & Ladle, R. J. (2008). Citations: Poor practices by authors reduce their value. *Nature*, *451*(7176), 244. <https://doi.org/10.1038/451244b>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Xie, J., Gong, K., Cheng, Y., & Ke, Q. (2019). The correlation between paper length and citations: A meta-analysis. *Scientometrics*, *118*(3), 763–786. <https://doi.org/10.1007/s11192-019-03015-0>
- Xie, J., Lu, H., Kang, L., & Cheng, Y. (2022). Citing criteria and its effects on researcher's intention to cite: A mixed-method study. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24614>