Measuring the effect of reviewers on manuscript change: a study on a sample of submissions to Royal Society journals (2006-2017)

Federico Bianchi[1], Daniel García-Costa[2], Francisco Grimaldo[2], and Flaminio Squazzoni*[1]

[1]*Department of Social and Political Sciences, University of Milan, Via Conservatorio 7, 20122 Milan, Italy*
[2]*Department of Computer Science, University of Valencia, Avinguda de la Universitat s/n, 46100 Burjassot, Spain*

This is the manuscript accepted for publication by the *Journal of Informetrics*. Please cite as follows:
Bianchi, F., García-Costa, D., Grimaldo, F., & Squazzoni, F. (2022). Measuring the effect of reviewers on manuscript change: a study on a sample of submissions to Royal Society journals (2006-2017). *Journal of Informetrics*, *16*(3): 101316. doi: `https://doi.org/10.1016/j.joi.2022.101316`

Abstract

Peer review is key for public trust of academic journals. It ensures that only rigorous research is published but also helps authors to increase the value of their manuscripts through feedback from reviewers. However, measuring the developmental value of peer review is difficult as it requires fine-grained manuscript data on various stages of the editorial process, which are rarely available. To fill this gap, we accessed complete data from Royal Society journals from 2006 to 2017, and measured manuscript changes during peer review from their initial submissions. We then estimated the effect of the number of reviewers and the evaluation of reviewers on manuscript development and their citations after publication. We found that the number of reviewers had an almost linear effect on manuscript change although with decreasing marginal effects whenever more than two reviewers were involved. This effect did not depend on the initial quality of manuscripts. We also found that changes due to reviewers tended to increase a manuscript's probability of being cited at least once after publication. While our findings show the multiple functions of peer review for manuscript development, research with larger and more representative journal samples is needed to develop more systematic measures that reflect the complexity of peer review.

*`flaminio.squazzoni@unimi.it`

1 Introduction

The digital age has witnessed an explosion of the means of scientific dissemination (Tennant et al., 2017). The proliferation of preprints, websites and online repositories has contributed to enhance the curation function of academic journals for scientific records (Squazzoni et al., 2020). The fact that we consider journals as synonymous of the quality of scientific records depends on the rigour of their internal evaluation standards and their capacity of adding value to submitted manuscripts (Baldwin, 2018). These standards can be achieved only if journals ensure rigorous selection of manuscripts and improve them through intensive collaboration between authors, reviewers and editors (Bornmann, 2011). Indeed, collaboration between editors, board members and external experts has greatly varied over time. This in turn has ensured that manuscript quality-screening and improvements has always been an intrinsic part of peer review at least since the 1950s in many research areas (Fyfe, Squazzoni, Torny, & Dondio, 2020; Merriman, 2020; Moxham & Fyfe, 2018).

Understanding whether and how these activities are performed by journals requires the examination of a variety of complex factors (Publons, 2018). Screening manuscripts and weeding out low-quality research requires the involvement of reviewers and editors, who reflect the best standards of research (Siler, Lee, & Bero, 2015). Developing manuscripts depends on a journal's capacity to create contexts within which a constructive dialogue between reviewers and authors is both fair and disinterested (Dondio, Casnici, Grimaldo, Gilbert, & Squazzoni, 2019).

Unfortunately, examining these factors jointly and empirically is difficult for various reasons, the most significant of which is the lack of fine-grained data from journals. For instance, research on peer review reports from repositories, such as Publons, helps to identify certain socio-demographic characteristics of reviewers and the choice of journals for which scholars typically review (Severin, Strinzel, Egger, Domingo, & Barros, 2021) or the connection between peer review activities and research productivity (Ortega, 20217). Recent research on a sample of peer review reports from Elsevier journals reconstructed the linguistic characteristics of reports depending on the type of recommendations and certain reviewer characteristics (Buljan, Garcia-Costa, Grimaldo, Squazzoni, & Marušić, 2020). Similarly, a recent study on a large-sample of reports from Elsevier journals found interesting heterogeneity in standards of reports depending on reviewer characteristics and areas of research (Garcia-Costa, Squazzoni, Mehmani, & Grimaldo, 2022). However, interlinking reports and manuscripts is impossible with a peer review report database, thus undermining the possibility of gauging the effect of peer review on manuscripts and on the journals themselves.

Research on the screening function of peer review typically concentrates on the reviewers' capability of predicting the quality of manuscripts (Casnici, Grimaldo, Gilbert, & Squazzoni, 2017). It has generally used ex-post measurements as an indirect proxy of reviewer reliability, including citations of different versions of manuscripts, e.g., published articles vs. rejected manuscripts later published in other journals, as well as differences in the impact factors of journals rejecting/publishing different versions of the same manuscripts (Rigby, Cox, & Julian, 2018). Unfortunately, only rarely have these studies included data on peer review reports and tracked manuscript change within the editorial process.

We believe that this is key to assess the developmental value of peer review as it allows us to examine how manuscripts change throughout the process of peer review (Atjonen, 2019; Bedeian, 2004; Matsui, Chen, Wang, & Ferrara, 2021; Rigby et al., 2018; Teplitskiy, 2016). For instance, the tendency of reducing the curation function of peer review to the goal of identifying impactful manuscripts via post-publication indicators (e.g., altmetrics, citations and other indicators), does not help to assess the quality of internal journal processes (Pontille & Torny, 2015; Seeber, 2020). However, without measuring how and how much manuscripts change throughout the process due to reviewer feedback, it is impossible to understand whether peer review adds anything relevant to the final manuscripts (Cowley, 2015).

Research examining these factors jointly is also essential to understand how journals harmonise different peer review functions for the benefit of their various stakeholders. The mechanics of peer review implies at least a triadic relationship with various expectations (Lugosi, 2021). Editors rely on reviewers to avoid publishing manuscripts of low quality and defend the prestige and position of their journals in a competitive, continually evolving environment (Liu, Hu, Wang, & Shi, 2018; Taşkın, Doğan, Kulczycki, & Zuccala, 2021). Authors expect that reviewers share constructive feedback for manuscript improvements, even when their manuscript is eventually rejected (Huisman & Smits, 2017). Reviewers expect authors to consider their comments and suggestions seriously to avoid being exploited while enforcing the highest scientific standards (S. Horbach & Halffman, 2018). The biases and inefficiencies of peer review are presently under the spotlight (Squazzoni et al., 2021; Tomkins, Zhang, & Heavlin, 2017) and many publishers are exploring innovative models to increase the transparency and accountability of the process, e.g., open peer or post-publication peer review, which require careful assessment (Eyre-Walker & Stoletzki, 2013; Harms & Credé, 2020; Thelwall, Allen, Papas, Nyakoojo, & Weigert, 2021). Thus, understanding manuscript change during peer review with data from multiple journals – and not only from individual cases (Grimaldo, Marušić, & Squazzoni, 2018) – can help us evaluate the importance of this fundamental academic institution more systematically (S. P. J. M. Horbach, 2021; Tennant & Ross-Hellauer, 2020).

Our paper aims to contribute to empirical research on peer review by presenting an explorative measurement of the developmental function of peer review. While previous research has investigated only specific journals and only rarely with complete data on manuscripts from each stage of the editorial process (Matsui et al., 2021; Teplitskiy, 2016), here, we have tested manuscript change during the editorial process with a large-scale, across-journal dataset and estimated possible effects on article citations. We aimed to test the effect of the number of reviewers and their evaluation on manuscript change within the editorial process and on later citations.

For this study, we first signed a confidential data sharing agreement with The Royal Society (Squazzoni, Grimaldo, & Marusic, 2017), the world's oldest independent scientific academy. The Royal Society pioneered the concepts and practices of academic journals, editorial responsibility and peer review (Fyfe, McDougall-Waters, & Moxham, 2015). Their journals include 11 titles, including *Philosophical Transactions A* and *Proceedings A*, which publish research on physical, mathematical and engineering sciences, *Philosophical Transactions B*, *Proceedings B* and *Biology Letters*, with a readership in biological sciences, as well as cross-disciplinary outlets, such as *Interface*, for cross-disciplinary research at the interface between the physical and life sciences, and *Royal Society Open Science*, the Royal Society's most recent open access journal in science, engineering and mathematics.

This agreement permitted us to collect complete and fully comparable temporal data on their journals from 2006 to 2017, including more than 10,000 manuscripts (see Methods). In order to ensure full comparability in terms of type of manuscripts and journals, we excluded all manuscripts submitted to the following four journals: *Open Biology*, *Interface Focus*, *Notes and Records* and *Biographical Memoirs*. Manuscripts from these journals were only weakly comparable with the rest of the sample, being mostly commentaries, short notes or reviews rather than research articles. We also restricted our sample to research articles, thus excluding any comments, reviews or notes.

After transforming all manuscript and review files of various format into text files, we calculated the Levenshtein distance (Levenshtein, 1966) between different versions of manuscripts to track any changes occurring throughout the process. Following Bravo, Farjam, Grimaldo, Birukou, and Squazzoni (2018), we built a *review score* that measured reviewer recommendations for each manuscript consistently, regardless of the different number of reviewers and rounds of reviews per manuscript. We considered this as a proxy of the initial quality of manuscripts as perceived by reviewers. We also calculated citations of published manuscripts to check whether changes during peer review could increase an article's probability of being cited after publication.

2 Methods

Our dataset included 10,996 manuscripts submitted to seven journals from the Royal Society from 2006 to 2017. Data included complete information regarding initial and revised versions of each submitted manuscript, including full text, reviewers' recommendations and editorial decisions.

In order to quantify the length of each manuscript, we converted each document into plain text files using dedicated *Python* libraries (i.e., '*docx*' for .doc and .docx files and '*pdfminer*' for .pdf files). We removed tables, figures, marks, rare characters, page headers and footers, as well as any irrelevant marks caused by document conversion. We then removed all non-ASCII characters. We downloaded the final version of all published articles from the Royal Society website. In the case of published articles, we divided their text into different portions and excluded images, figures and tables, thus standardizing their format with their related submission files. This allowed us to assign a unique ID to different files of the same manuscript (e.g., original submissions and published articles).

We measured the *text changes* by computing the difference between the originally submitted manuscript and either the published or the rejected version. We computed the Levenshtein distance (Levenshtein, 1966) between different text versions, i.e., the number of changes needed to convert one text string into another, thus detecting any change of the text throughout the various stages of peer review. We preferred this measurement to token-based distances, such as cosine or Jaccard distance, as the latter would not have permitted us to consider certain changes, such as the syntax or rephrasing using the same words.

When calculating *text changes* with the Levenshtein distance, we also calculated the difference between the originally submitted manuscript and the final version (either the published article or the rejected manuscript) in their listed references. In order to identify references, we used various regular expressions (*regex*) which were shared by different referencing styles (e.g., IEEE, Vancouver, APA). We defined the regex to extract separately the publication year, the title and the list of authors. We then calculated a similarity ratio that considered two references as equal when: (i) both sources reported the same publication year; (ii) the cosine distance between titles was smaller than 0.1; and (iii) either both references had the same number of authors or the cosine distance between the list of authors was smaller than 0.1. We set this threshold to 0.1 after manual experimentation on the data. We used the cosine distance as any token-based distances was less sensitive to small spelling changes when comparing references.

We calculated the *reference changes* as follows:

$$1 - \frac{\textit{Number of similar references}}{\textit{Max number of references in either documents}}.$$

For the sake of interpretation, we re-scaled both *text changes* and *reference changes* to a 0-100 range.

We then calculated the *number of reviewers* for each manuscript by counting the total number of reviewers involved in all rounds of reviews. For instance, assume that in the first round, a manuscript was reviewed by reviewers 1 and 2 and that in the second round, reviewer 2 was not involved while the editor contacted reviewer 3. In these cases, we counted a total number of reviewers = 3. This was to reflect the fact that a manuscript can change due to the effect of each individual reviewer to whom it was exposed.

By following Bravo et al. (2018), we calculated a *review score* for each manuscript as a proxy of the manuscript's quality resulting from reviewers' recommendations. This score allowed us to compare the evaluation of manuscripts submitted to various journals regardless of differences in the number of reviewers per manuscript. We built a set of all possible unique combinations of recommendations for each manuscript (e.g., *{accept, accept}*, *{accept, minor revision}*, *{accept, major revision}*, ..., *{reject, reject}*) and counted the number of combinations that were less favourable (*# worse*) or more favourable (*# better*) than the recommendation received by the manuscript (e.g., *{accept, accept}* was better than *{accept, major revision}*). We handled combinations which could not be considered as clearly better or worse as reported in Bravo et al. (2018, Table 2). After testing all possible (better or worse) combinations per manuscript and verifying lack of differences on the outcomes, we calculated the review scores as follows:

$$\textit{review score} = \frac{\#\ \textit{worse}}{\#\ \textit{worse} + \#\ \textit{better}}.$$

We measured inter-reviewer *agreement* by calculating the number of similar recommendations divided by the total number of reviews per manuscript at the first round (e.g., 2/3 agreement in case of three reviewers recommending *{minor revision, major revision, major revision}*). Finally, we measured the *impact* of published manuscripts by calculating the number of citations for each article using the DOI obtained from the Royal Society journal platform to query Altmetrics API on Dimensions.ai database (Khan, Arjmandi, & Yuvaraj, 2021).

3 Results

The length of the text of originally submitted manuscripts was highly left-skewed. The median length was 21,773 characters. Figure 1 shows that the final version of both published and rejected manuscripts changed considerably in terms of Levenshtein distance compared to their initial version. This was true for both *text changes* (M = 40.72 %, SD = 15.67 %) and *reference changes* (M = 41.33 %, SD = 21.42 %). Most manuscripts were reviewed by at least two reviewers (65.60 %), with only a minority reviewed by three or more reviewers.

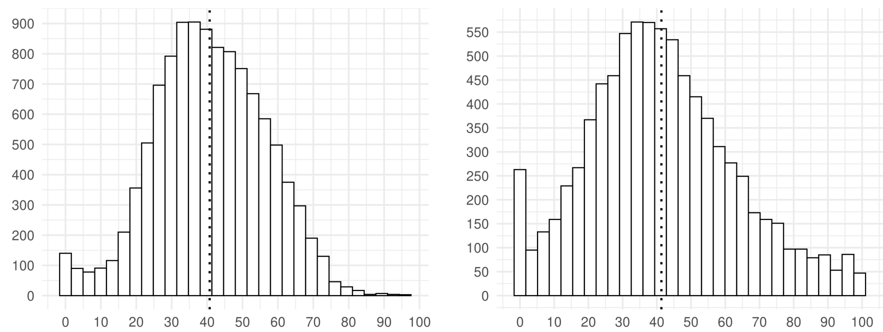

Figure 1: Distribution of *text changes* (left) and *reference changes* (right) among sampled manuscripts, measured by the Levenshtein distance (%) between the original submission and the final version (either published articles or rejected manuscripts). Vertical dashed lines indicate mean values.

We tested the impact of the *number of reviewers* on manuscript change by estimating linear mixed-effect models with random intercepts on *text changes* and *reference changes*. With regard to the former, results showed that the number of reviewers tended to increase manuscript changes (see Figure 2a). Changes increased almost linearly with the number of reviewers. However, a greater effect was found when shifting from one to two reviewers evaluating the same manuscript in various rounds of the process. Note that whenever manuscripts were evaluated by five or more reviewers, we found decreasing marginal effects compared to the case of manuscripts evaluated by four reviewers. We found a similarly positive effect on *reference changes* when manuscripts were assessed by up to four reviewers. Note that this effect decreased whenever manuscripts were assessed by more than four reviewers (see Figure 2b). In both models, the effect of the *number of reviewers* was estimated by controlling for journal-specific heterogeneity (random intercepts), the length of the originally submitted manuscripts, the *review score*, i.e., the quality of manuscripts in reviewers' opinion, and the inter-reviewer agreement (see Table 1).

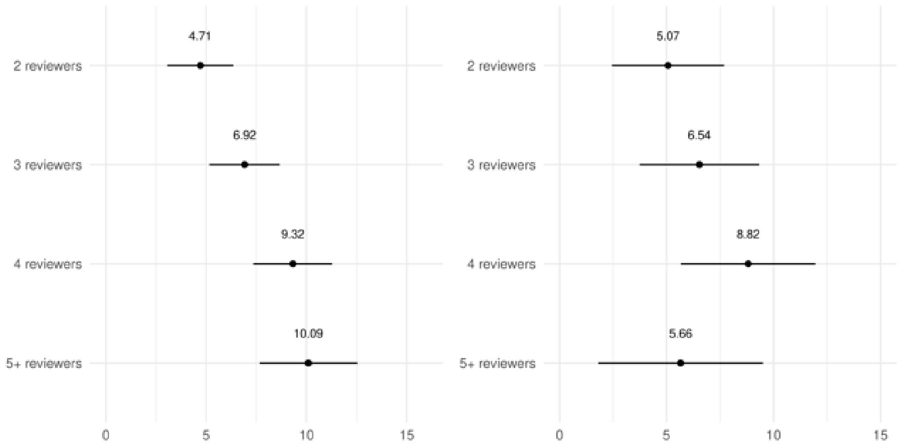

Figure 2: Linear mixed-effects models of *text changes* (left) and *reference changes* (right), measured through Levenshtein distance (%): Estimated fixed effects of *number of reviewers* (*dots*, reference category: "1 reviewer") with 95% confidence intervals (*lines*)). The models include all control variables presented in Table 1 and random intercepts of different journals.

	text changes				reference changes			
	$\hat{\beta}$	S.E.	95% C.I.	p	$\hat{\beta}$	S.E.	95% C.I.	p
Fixed effects								
2 reviewers	4.71	0.84	[3.06, 6.36]	0.00	5.07	1.34	[2.45, 7.70]	0.00
3 reviewers	6.92	0.89	[5.16, 8.67]	0.00	6.54	1.43	[3.74, 9.34]	0.00
4 reviewers	9.32	1.00	[7.36, 11.29]	0.00	8.82	1.60	[5.67, 11.96]	0.00
5+ reviewers	10.09	1.24	[7.66, 12.53]	0.00	5.66	1.96	[1.81, 9.51]	0.00
Length of original submission	0.00	0.00	[0.00, 0.00]	0.01	0.00	0.00	[0.00, 0.00]	0.03
Review score	0.04	0.01	[0.03, 0.06]	0.00	0.11	0.01	[0.09, 0.13]	0.00
Reviewer agreement	-0.06	0.01	[-0.07, -0.05]	0.00	-0.04	0.01	[-0.06, -0.02]	0.00
Constant	37.78	2.10	[33.67, 41.90]	0.00	33.83	2.77	[28.40, 39.26]	0.00
Random effects								
SD (Intercept)	4.56				5.19			
Number of observations	10,308				7,777			

Table 1: Linear mixed-effects models estimating the effect of the *number of reviewers* (reference category: “1 reviewer") on *text changes* and *reference changes* with journal-specific random intercepts. Note that the number of observations varied due to cases of manuscript files without correctly formatted or reported references.

With regard to the effect of manuscript change on published articles' *impact*, Figure 3 shows that the distribution of manuscripts cited at least one time after being published was relatively heterogeneous across the journals. Overall, the average number of citations was 22.64 (SD = 39.82).

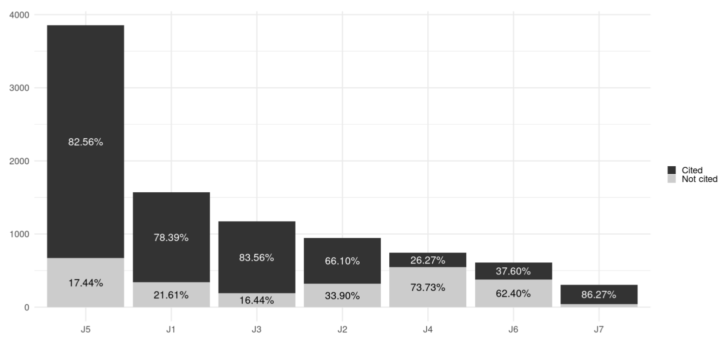

Figure 3: Distribution of published articles which had received at least one citation (dark) vs. those which had not received any citation (light) across journals. Values reported inside bar relate to within-journal percentages.

Tables 2 and 3 show two logistic regression models estimating a small positive effect of *text changes* and *reference changes* respectively on *impact*. In both models, we controlled for differences between journals, which significantly varied in terms of impact factor, and time exposure of articles, which could affect citation dynamics. Note that the distribution of the number of citations was highly skewed (5.39), thus making linear regression models poorly informative. This led us to consider a binary variable, i.e., whether articles had received at least one citation or not. We also estimated zero-inflated negative binomial regression models (Hilbe, 2014), which suggested that evidence of a small effect of *text* and *reference changes* could be found only in changing between receiving no citations or receiving at least one (see Additional analysis for more details), adjusting for across-journal differences and years from publication. However, note that estimating the effect of changes due to peer review on citations is problematic because of other possible confounding factors, including authors' reputation or particular characteristics of the published study (e.g., the popularity of the topic).

	Odds	C.I. 95%	S.E.	p
Text changes	1.01	[1.01, 1.02]	(0.00)	0.00
Review score	1.00	[1.00, 1.01]	(0.00)	0.08
Journal 2	0.32	[0.25, 0.40]	(0.12)	0.00
Journal 3	0.28	[0.22, 0.36]	(0.13)	0.00
Journal 4	0.04	[0.03, 0.05]	(0.14)	0.00
Journal 5	1.29	[1.09, 1.52]	(0.08)	0.00
Journal 6	0.03	[0.02, 0.04]	(0.16)	0.00
Journal 7	0.59	[0.33, 1.15]	(0.31)	0.09
Years published	0.69	[0.68, 0.71]	(0.01)	0.00
Constant	25.94	[17.93, 37.67]	(0.19)	0.00
Number of observations		8,589		
Log likelihood		-3429.04		

Table 2: Logistic regression model estimating the effect of *text changes* on an article's probability of being cited at least once after publication. (Reference category of *journal*: "Journal 1")

	Odds	C.I. 95%	S.E.	p
Reference changes	1.01	[1.01, 1.01]	(0.00)	0.00
Review score	1.00	[1.00, 1.00]	(0.00)	0.32
Journal 2	0.32	[0.25, 0.42]	(0.13)	0.00
Journal 3	0.41	[0.31, 0.55]	(0.15)	0.00
Journal 4	0.04	[0.03, 0.06]	(0.02)	0.00
Journal 5	1.43	[1.20, 1.70]	(0.09)	0.00
Journal 6	0.04	[0.03, 0.06]	(0.02)	0.00
Journal 7	0.67	[0.36, 1.35]	(0.33)	0.22
Years published	0.71	[0.69, 0.72]	(0.01)	0.00
Constant	24.42	[17.02, 35.23]	(0.19)	0.00
Number of observations		6,653		
Log likelihood		-2844.44		

Table 3: Logistic regression model estimating the effect of *reference changes* on an article's probability of being cited at least once after publication. (Reference category of *journal*: "Journal 1")

4 Discussion and conclusions

The credibility of academic journals greatly depends on the quality of peer review (Bornmann, 2011; Edwards & Siddhartha, 2017; Kharasch et al., 2021). Screening manuscripts without providing constructive feedback to authors to help them improving their manuscripts is not a good practice, especially whenever journals must ensure that only rigorous science is published (Atjonen, 2019; Teplitskiy, 2016). Although this may come at the price of delaying publications, constructive and elaborated peer review is also key for expert learning (Rigby et al., 2018).

Our study contributes to research on the developmental function of peer review (Atjonen, 2019; Garcia-Costa et al., 2022; Matsui et al., 2021; Seeber, 2020; Strang & Siler, 2015) by exploring a large dataset of manuscripts, editorial decisions and peer review outcomes from journals from the Royal Society. Our results showed that reviewers had a considerable impact on manuscript changes. Exposing manuscripts to reviewer evaluations in various peer review rounds led to an average level of about 40% of changes in manuscript text and references. Manuscript change tended to increase with the number of reviewers assessing the same manuscript and this effect was independent of the initial quality of manuscripts. Not only were manuscripts of moderate initial quality improved during peer review, but also manuscripts initially receiving more positive evaluations from reviewers, as well as those determining lowest inter-reviewer agreement, were refined and changed throughout the process. Furthermore, this effect was found regardless of any journal specificity.

Unfortunately, our analysis could not focus on details on the content of reviewer requests. While reference changes would indicate that reviewers requested authors to add relevant literature, only a linguistic analysis of the content of reports could help us to disentangle requests for conceptual developments or methodological improvements. A comprehensive analysis would also require us to match requests by reviewers and revisions made by authors, which could be made only by reducing the sample size at the expense of generalisation (Eve et al., 2021).

With all due caveats regarding possible confounding factors, we found that manuscript changes increased the probability that a published article was cited at least once after publication. However, this finding should be considered with caution. Previous research has showed mixed evidence on the link between peer review and article citations, suggesting that reviewers do not systematically predict the future impact of articles in terms of citations (Teplitskiy, 2016). The essential element of developmental peer review is to help authors improve their manuscripts, whereas the impact of articles depends on various factors (Coupé, 2013). For instance, it is difficult to estimate whether citations of manuscripts are related to the quality of manuscripts as outcome of peer review, the reputation of authors or to interest in the manuscripts' topics (Seeber, 2020). Here, future research on the developmental function of peer review should consider these complex factors more systematically, although fine-grained data required to study these aspects are rarely available, e.g., the integration of journal data with scientific records of authors prior to submitting their manuscripts (Squazzoni et al., 2020).

Finally, as suggested by a recent systematic review on experimental interventions (Gaudino et al., 2021), improving the developmental function of peer review calls for problems of sustainability and publication time delay (Merrill, 2014). There is a clear trade-off between peer review functions, including quality and efficient use of reviewer time (Bianchi, Grimaldo, Bravo, & Squazzoni, 2018). Unfortunately, there is still scant knowledge on these multiple functions of peer review, including the effect of reviewer guidelines (or lack of), the role ambiguity of editors and reviewers with often unclear editorial decision-making responsibility (Seeber, 2020; Song et al., 2021; Tennant & Ross-Hellauer, 2020). More research is needed to assess these trade-offs and examine the effect of peer review on the quality and recognition of manuscripts. This will mostly depend on our collective capability of removing obstacles of data sharing between publishers, journals and the scientific community (Squazzoni et al., 2020).

Data accessibility

The dataset for full replication of our study is provided here: `https://dataverse.harvard.edu/privateurl.xhtml?token=6bde093d-dc44-4702-92ef-741f2e166e83`. As mentioned in the text, data access required a confidentiality agreement to be signed with the Royal Society, which included journal anonymization.

Competing interests

The authors declare we have no competing interests.

Funding

DG-C and FG were partially supported by the Spanish Ministry of Science, Innovation and Universities (MCIU), the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERDF) under project RTI2018-095820-B-I00. FB was supported by a PRIN-MUR (Progetti di Rilevante Interesse Nazionale – Italian Ministry of University and Research) grant (Grant Number: 20178TRM3F001 "14All"). FS was supported by a Transition Grant from the University of Milan (PSR2015-17).

Acknowledgements

We gratefully acknowledge Phil Hurst and the team of The Royal Society for providing data and covering the cost of their extraction from manuscript submission systems. We would like to thank the journal reviewers for helpful comments.

Appendix: Additional analysis

Figure 4 (left) shows the distribution of the number of citations received by published articles. The average number of citations was 22.87 (SD = 39.93), while the median number was 11. The right side of Figure 4 shows the distribution of the log-transformed number of citations, according to $\ln(\textit{number of citations}+1)$. A Kolmogorov-Smirnov test of normality reported strong evidence against the log-linearity of the distribution of the number of citations ($D = 0.62$, $p = 0.00$).

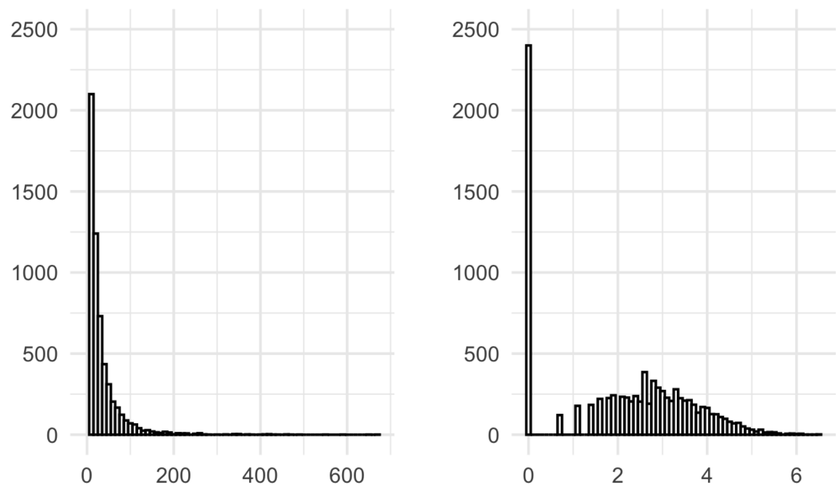

Figure 4: Distribution of the number of citations (left) and the logarithmic transformation (right) among published articles.

Figure 5 shows the number of published articles with zero citations (25.30%) compared to the number of articles which were cited at least once.

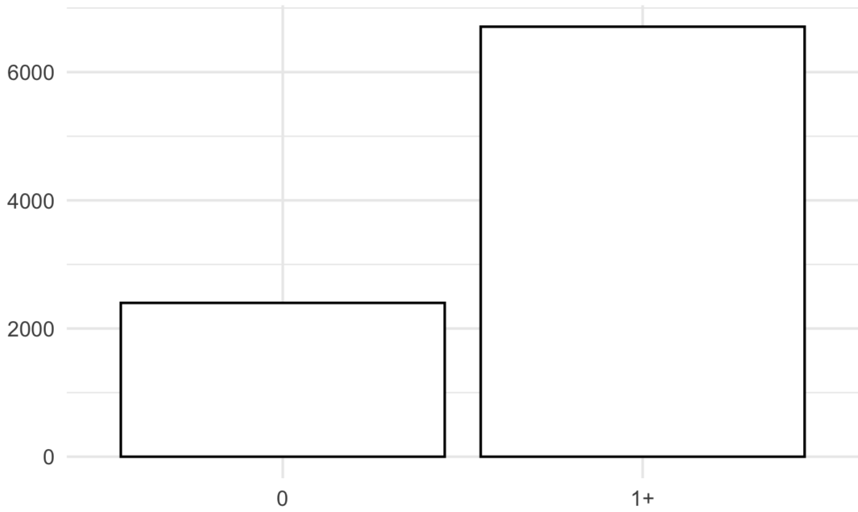

Figure 5: Number of published articles with zero vs. at least one citations.

Table 4 and Table 5 show the estimates of Zero-Inflated Negative Binomial (ZINB) regression models (Hilbe, 2014) in which the number of article citations was considered as a function of the same set of regressors reported in Tables 2 and 3, respectively. ZINB regressions consider the binary event of scoring 0 (zero-inflation model) separately from the count scores of an outcome (count model). The reported models show that both *text* and *reference changes* implied a small negative effect on the probability of receiving 0 citations against those of receiving at least one. With regard to the count models, we did not find any evidence of an effect of *text changes*, while we found a null effect of *reference changes*.

	Count model			Zero-inflation model		
	$\hat{\beta}$	S.E.	$\Pr(>\|z\|)$	$\hat{\beta}$	S.E.	$\Pr(>\|z\|)$
Text changes	0.00	0.00	0.32	-0.01	0.00	0.00
Review score	0.00	0.00	0.00	0.00	0.00	0.12
Journal 2	0.08	0.04	0.06	1.10	0.12	0.00
Journal 3	-0.08	0.04	0.03	1.11	0.15	0.00
Journal 4	-0.19	0.07	0.00	3.23	0.15	0.00
Journal 5	0.39	0.03	0.00	-0.26	0.09	0.00
Journal 6	-0.11	0.08	0.17	3.53	0.17	0.00
Journal 7	0.41	0.06	0.00	0.32	0.45	0.48
Years published	0.19	0.00	0.00	0.39	0.01	0.00
Constant	2.00	0.06	0.00	-3.49	0.20	0.00
Log(θ)	0.51	0.02	0.00			

Table 4: Zero-Inflation Negative Binomial Regression model of the number of citations as a function of *text changes* and the same covariates as in Table 2.

	Count model			Zero-inflation model		
	$\hat{\beta}$	S.E.	Pr(> $\|z\|$)	$\hat{\beta}$	S.E.	Pr(> $\|z\|$)
Reference changes	0.00	0.00	0.00	-0.01	0.00	0.00
Review score	0.00	0.00	0.00	0.00	0.00	0.44
Journal 2	1.11	0.05	0.5	1.11	0.14	0.00
Journal 3	-0.05	0.05	0.29	0.67	0.18	0.00
Journal 4	0.25	0.09	0.01	3.17	0.18	0.00
Journal 5	0.45	0.03	0.00	-0.35	0.09	0.00
Journal 6	0.00	0.10	0.98	3.39	0.20	0.00
Journal 7	0.44	0.08	0.00	0.27	0.44	0.53
Years published	0.19	0.00	0.00	0.37	0.01	0.00
Constant	2.00	0.06	0.00	-3.49	0.20	0.00
Log(θ)	0.54	0.02	0.00			

Table 5: Zero-Inflation Negative Binomial Regression model of the number of citations as a function of *reference changes* and the same covariates as in Table 3.

Furthermore, we modelled the number of citations as a 4-level ordinal variable based on quartiles. Tables 6 and 7 show results from ordinal logistic regression models (McCullagh, 1980) as a function of the same set of regressors reported in Tables 2 and 3, respectively. In both models, we found a small effect of *text* and *reference changes*.

	$\hat{\beta}$	S.E.	Pr(> $\|t\|$)
Text changes	0.01	0.00	0.00
Review score	0.00	0.00	0.00
Journal 2	-0.59	0.08	0.00
Journal 3	-0.55	0.07	0.00
Journal 4	-2.36	0.10	0.00
Journal 5	0.62	0.06	0.00
Journal 6	-1.67	0.13	0.00
Journal 7	0.08	0.13	0.52
Years published	-0.04	0.01	0.00
1|2	-1.08	0.12	0.00
2|3	0.25	0.12	0.03
3|4	1.46	0.12)	0.00

Table 6: Ordinal logistic regression model of quartiles of number of citations as a function of *text changes* and the same covariates as in Table 3.

	$\hat{\beta}$	S.E.	Pr(> |t|)
Reference changes	0.01	0.00	0.00
Review score	0.00	0.00	0.00
Journal 2	-0.69	0.10	0.00
Journal 3	-0.39	0.08	0.00
Journal 4	-2.56	0.13	0.00
Journal 5	0.63	0.07	0.00
Journal 6	-1.56	0.16)	0.00
Journal 7	0.14	0.15	0.33
Years published	-0.04	0.01	0.00
1|2	-0.81	0.12	0.00
2|3	0.33	0.12	0.00
3|4	1.46	0.12	0.00

Table 7: Ordinal logistic regression model of quartiles of number of citations as a function of *reference changes* and the same covariates as in Table 3.

References

Atjonen, P. (2019). Peer review in the development of academic articles: Experiences of finnish authors in the educational sciences. *Learned Publishing*, *32*(2), 137-146. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/leap.1204` doi: https://doi.org/10.1002/leap.1204

Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of "peer review" in the cold war united states. *Isis*, *109*(3), 538-558. Retrieved from `https://doi.org/10.1086/700070` doi: 10.1086/700070

Bedeian, A. G. (2004). Peer review and the social construction of knowledge in the management discipline. *Academy of Management Learning & Education*, *3*(2), 198-216. Retrieved from `https://doi.org/10.5465/amle.2004.13500489` doi: 10.5465/amle.2004.13500489

Bianchi, F., Grimaldo, F., Bravo, G., & Squazzoni, F. (2018). The peer review game: an agent-based model of scientists facing resource constraints and institutional pressures. *Scientometrics*, *116*, 1401–1420. doi: 10.1007/s11192-018-2825-4

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, *45*(1), 197-245. Retrieved from `https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.2011.1440450112` doi: 10.1002/aris.2011.1440450112

Bravo, G., Farjam, M., Grimaldo, F., Birukou, A., & Squazzoni, F. (2018). Hidden connections: network effects on editorial decisions in four computer science journals. *Journal of Informetrics*, *12*(1), 101-112. doi: 10.1016/j.joi.2017.12.002

Buljan, I., Garcia-Costa, D., Grimaldo, F., Squazzoni, F., & Marušić, A. (2020, jul). Meta-research: Large-scale language analysis of peer review reports. *eLife*, *9*, e53249. Retrieved from https://doi.org/10.7554/eLife.53249 doi: 10.7554/eLife.53249

Casnici, N., Grimaldo, F., Gilbert, N., & Squazzoni, F. (2017). Attitudes of referees in a multidisciplinary journal: An empirical analysis. *Journal of the Association for Information Science and Technology*, *68*(7), 1763–1771. doi: 10.1002/asi.23665

Coupé, T. (2013). Peer review versus citations – an analysis of best paper prizes. *Research Policy*, *42*(1), 295-301. Retrieved from https://www.sciencedirect.com/science/article/pii/S0048733312001412 doi: https://doi.org/10.1016/j.respol.2012.05.004

Cowley, S. J. (2015). How peer-review constrains cognition: on the frontline in the knowledge sector. *Frontiers in Psychology*, *6*, 1706. doi: 10.3389/fpsyg.2015.01706

Dondio, P., Casnici, N., Grimaldo, F., Gilbert, N., & Squazzoni, F. (2019). The "invisible hand" of peer review: The implications of author-referee networks on peer review in a scholarly journal. *Journal of Informetrics*, *13*(2), 708-716. Retrieved from https://www.sciencedirect.com/science/article/pii/S1751157718304206 doi: https://doi.org/10.1016/j.joi.2019.03.018

Edwards, M. A., & Siddhartha, R. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, *34*(1), 51-61. doi: 10.1089/ees.2016.0223

Eve, M. P., Neylon, C., O'Donnell, D. P., Moore, S., Gadie, R., Odeniyi, V., & S., P. (2021). *Reading peer review. plos one and institutional change in academia.* Cambridge University Press.

Eyre-Walker, A., & Stoletzki, N. (2013, 10). The assessment of science: The relative merits of post-publication review, the impact factor, and the number of citations. *PLOS Biology*, *11*(10), 1-8. Retrieved from https://doi.org/10.1371/journal.pbio.1001675 doi: 10.1371/journal.pbio.1001675

Fyfe, A., McDougall-Waters, J., & Moxham, N. (2015). 350 years of scientific periodicals. *Notes and Records: the Royal Society Journal of the History of Science*, *69*(3), 227-239. doi: 10.1098/rsnr.2015.0036

Fyfe, A., Squazzoni, F., Torny, D., & Dondio, P. (2020). Managing the growth of peer review at the royal society journals, 1865-1965. *Science, Technology, & Human Values*, *45*(3), 405-429. Retrieved from https://doi.org/10.1177/0162243919862868 doi: 10.1177/0162243919862868

Garcia-Costa, D., Squazzoni, F., Mehmani, B., & Grimaldo, F. (2022). Measuring the developmental function of peer review: a multi-dimensional, cross-disciplinary analysis of peer review reports from 740 academic journals. *PeerJ*, *10*(e13539). doi: https://doi.org/10.7717/peerj.13539

Gaudino, M., Robinson, N. B., Franco, A. D., Hameed, I., Naik, A., Demetres, M., ... Biondi-Zoccai, G. (2021). Effects of experimental interventions to improve the biomedical peer-review process: A systematic review and meta-analysis. *Journal of the American Heart Association*, *10*(15), e019903. doi: 10.1161/JAHA.120.019903

Grimaldo, F., Marušić, A., & Squazzoni, F. (2018, 02). Fragments of peer review: A quantitative analysis of the literature (1969-2015). *PLOS ONE*, *13*(2), 1-14. Retrieved from `https://doi.org/10.1371/journal.pone.0193148` doi: 10.1371/journal.pone.0193148

Harms, P. D., & Credé, M. (2020). Bringing the review process into the 21st century: Post-publication peer review. *Industrial and Organizational Psychology*, *13*(1), 51–53. doi: 10.1017/iop.2020.13

Hilbe, J. M. (2014). *Modeling count data*. New York, NY: Cambridge University Press.

Horbach, S., & Halffman, W. (2018). The changing forms and expectations of peer review. *Res Integr Peer Rev*, *3*(8). Retrieved from `https://doi.org/10.1186/s41073-018-0051-5` doi: 10.1186/s41073-018-0051-5

Horbach, S. P. J. M. (2021, 01). No time for that now! Qualitative changes in manuscript peer review during the Covid-19 pandemic. *Research Evaluation*. Retrieved from `https://doi.org/10.1093/reseval/rvaa037` (rvaa037) doi: 10.1093/reseval/rvaa037

Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: the author's perspective. *Scientometrics*, *113*, 633–650. Retrieved from `https://doi.org/10.1007/s11192-017-2310-5` doi: 10.1007/s11192-017-2310-5

Khan, D., Arjmandi, M. K., & Yuvaraj, M. (2021). Most cited works on cloud computing: The 'citation classics' as viewed through dimensions.ai. *Science & Technology Libraries*, *0*(0), 1-14. Retrieved from `https://doi.org/10.1080/0194262X.2021.1951424` doi: 10.1080/0194262X.2021.1951424

Kharasch, E. D., Avram, M. J., Clark, J. D., Davidson, A. J., Houle, T. T., Levy, J. H., ... Vutskits, L. (2021, 01). Peer review matters: Research quality and the public trust. *Anesthesiology*, *134*(1), 1-6. Retrieved from `https://doi.org/10.1097/ALN.0000000000003608` doi: 10.1097/ALN.0000000000003608

Levenshtein, V. I. (1966, feb). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, *10*(8), 707–710. (Doklady Akademii Nauk SSSR, V163 No4 845-848 1965)

Liu, M., Hu, X., Wang, Y., & Shi, D. (2018). Survive or perish: Investigating the life cycle of academic journals from 1950 to 2013 using survival analysis methods. *Journal of Informetrics*, *12*(1), 344-364. Retrieved from https://www.sciencedirect.com/science/article/pii/S175115771730281X doi: https://doi.org/10.1016/j.joi.2018.02.001

Lugosi, P. (2021). The value creation cycle of peer review. *Annals of Tourism Research*, *86*, 103092. Retrieved from https://www.sciencedirect.com/science/article/pii/S016073832030236X doi: https://doi.org/10.1016/j.annals.2020.103092

Matsui, A., Chen, E., Wang, Y., & Ferrara, E. (2021). The impact of peer review on the contribution potential of scientific papers. *PeerJ*, *9*, e11999. doi: 10.7717/peerj.11999

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B*, *42*(2), 109-142.

Merrill, E. (2014). Reviewer overload and what can we do about it. *The Journal of Wildlife Management*, *78*(6), 961-962. Retrieved from https://wildlife.onlinelibrary.wiley.com/doi/abs/10.1002/jwmg.763 doi: 10.1002/jwmg.763

Merriman, B. (2020). Peer review as an evolving response to organizational constraint: Evidence from sociology journals, 1952–2018. *The American Sociologist*, *52*, 341—366. doi: 10.1007/s12108-020-09473-x

Moxham, N., & Fyfe, A. (2018). The Royal Society and the prehistory of peer review, 1665–1965. *The Historical Journal*, *61*(4), 863–889. doi: 10.1017/S0018246X17000334

Ortega, J. (20217). Are peer-review activities related to reviewer bibliometric performance? a scientometric analysis of publons. *Scientometrics*, *112*, 947–962. doi: 10.1007/s11192-017-2399-6

Pontille, D., & Torny, D. (2015). From manuscript evaluation to article valuation: The changing technologies of journal peer review. *Human Studies*, *38*, 57-79. Retrieved from https://doi.org/10.1007/s10746-014-9335-z doi: 10.1007/s10746-014-9335-z

Publons. (2018). *2018 global state of peer review.* (Clarivate Analytics)

Rigby, J., Cox, D., & Julian, K. (2018). Journal peer review: a bar or bridge? An analysis of a paper's revision history and turnaround time, and the effect on citation. *Scientometrics*, *114*(3), 1087-1105. doi: 10.1007/s11192-017-2630-5

Seeber, M. (2020). How do journals of different rank instruct peer reviewers? Reviewer guidelines in the field of management. *Scientometrics*, *122*, 1387–140. Retrieved from https://doi.org/10.1007/s11192-019-03343-1 doi: 10.1007/s11192-019-03343-1

Severin, A., Strinzel, M., Egger, M., Domingo, M., & Barros, T. (2021). Characteristics of scholars who review for predatory and legitimate journals: linkage study of cabells scholarly analytics and publons data. *BMJ Open*, *11*(7). Retrieved from https://bmjopen.bmj.com/

content/11/7/e050270 doi: 10.1136/bmjopen-2021-050270

Siler, K., Lee, K., & Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences*, *112*(2), 360–365. Retrieved from https://www.pnas.org/content/112/2/360 doi: 10.1073/pnas.1418218112

Song, E., Ang, L., Park, J.-Y., Jun, E.-Y., Kim, K. H., Jun, J., ... Lee, M. S. (2021, 05). A scoping review on biomedical journal peer review guides for reviewers. *PLOS ONE*, *16*(5), 1-18. Retrieved from https://doi.org/10.1371/journal.pone.0251440 doi: 10.1371/journal.pone.0251440

Squazzoni, F., Ahrweiler, P., Barros, T., Bianchi, F., Birukou, A., Blom, H. J. J., ... Willis, M. (2020). Unlock ways to share data on peer review. *Nature*, *578*, 512-514. doi: 10.1038/d41586-020-00500-yz

Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., Willis, M., ... Grimaldo, F. (2021). Peer review and gender bias: A study on 145 scholarly journals. *Science Advances*, *7*(2), eabd0299. Retrieved from https://www.science.org/doi/abs/10.1126/sciadv.abd0299 doi: 10.1126/sciadv.abd0299

Squazzoni, F., Grimaldo, F., & Marusic, A. (2017). Publishing: Journals could share peer-review data. *Nature*, *546*(352). doi: 10.1038/546352a

Strang, D., & Siler, K. (2015). Revising as reframing: Original submissions versus published papers in administrative science quarterly, 2005 to 2009. *Sociological Theory*, *33*(1), 71-96. Retrieved from https://doi.org/10.1177/0735275115572152 doi: 10.1177/0735275115572152

Taşkın, Z., Doğan, G., Kulczycki, E., & Zuccala, A. A. (2021). Self-citation patterns of journals indexed in the journal citation reports. *Journal of Informetrics*, *15*(4), 101221. Retrieved from https://www.sciencedirect.com/science/article/pii/S1751157721000924 doi: https://doi.org/10.1016/j.joi.2021.101221

Tennant, J., Dugan, J., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., ... Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review [version 3; referees: 2 approved]. *F1000 Research*, *6*(1151). doi: 10.12688/f1000research.12037.3

Tennant, J., & Ross-Hellauer, T. (2020, 5). The limitations to our understanding of peer review. *Research Integrity & Peer Review*, *6*. doi: 10.1186/s41073-020-00092-1

Teplitskiy, M. (2016). Frame search and re-search: How quantitative sociological articles change during peer review. *The American Sociologist*, *47*, 264–288. Retrieved from https://doi.org/10.1007/s12108-015-9288-3 doi: 10.1177/0162243919862868

Thelwall, M., Allen, L., Papas, E.-R., Nyakoojo, Z., & Weigert, V. (2021). Does the use of open, non-anonymous peer review in scholarly publishing introduce bias? Evidence from the f1000research post-publication open peer review publishing model. *Journal of Information Science*, *47*(6), 809-820. Retrieved from `https://doi.org/10.1177/0165551520938678` doi: 10.1177/0165551520938678

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences*, *114*(48), 12708–12713. Retrieved from `https://www.pnas.org/content/114/48/12708` doi: 10.1073/pnas.1707323114