

# Accountability in Research

## Ethics, Integrity and Policy

ISSN: 0898-9621 (Print) 1545-5815 (Online) Journal homepage: [www.tandfonline.com/journals/gacr20](http://www.tandfonline.com/journals/gacr20)

## Gaming the peer review system: Evidence for a review mill in medicine highlights the need to ensure reviewer integrity

M. Ángeles Oviedo-García, René Aquarius & Dorothy V. M. Bishop

To cite this article: M. Ángeles Oviedo-García, René Aquarius & Dorothy V. M. Bishop (05 Mar 2026): Gaming the peer review system: Evidence for a review mill in medicine highlights the need to ensure reviewer integrity, *Accountability in Research*, DOI: [10.1080/08989621.2026.2640012](https://doi.org/10.1080/08989621.2026.2640012)

To link to this article: <https://doi.org/10.1080/08989621.2026.2640012>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 05 Mar 2026.



Submit your article to this journal [↗](#)



Article views: 298



View related articles [↗](#)



View Crossmark data [↗](#)

# Gaming the peer review system: Evidence for a review mill in medicine highlights the need to ensure reviewer integrity

M. Ángeles Oviedo-García <sup>a</sup>, René Aquarius <sup>b</sup> and Dorothy V. M. Bishop <sup>c</sup>

<sup>a</sup>Business Management and Marketing Department, University of Seville, Seville, Spain; <sup>b</sup>Department of Neurosurgery, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>c</sup>Department of Experimental Psychology, University of Oxford, Oxford, UK

## ABSTRACT

**Background:** Review mills are recognized when individuals generate numerous generic review reports, typically containing suggestions for citations to their own work. Here, we report a network with characteristics of a review mill in the field of gynecological oncology.

**Methods:** Our search started with a review that contained “boilerplate” comments as well as suggestions that specific PubMed IDs be cited. We searched the internet using Google for review reports using the same boilerplate comments. We coded text to quantify similarities between reviews and compiled citations suggested by reviewers. For comparison, we analyzed 59 reviews of the same articles by other peer reviewers.

**Results:** We identified a network of 195 review reports that shared boilerplate text from 170 articles. One hundred and eighty-six reports suggested citing articles coauthored by a member of the network. Five members of the network had editorial roles. Authors of 142 articles complied with suggestions for citation. Boilerplate text and citation recommendations were rare in the comparison reports.

**Conclusions:** Review mills lead to articles being published without proper peer review. This is of particular concern in medical research. Open peer review and transparent reporting of the editors responsible for handling papers will make it easier to detect review mills.

## ARTICLE HISTORY

Received 4 January 2026  
Accepted 27 February 2026.

## KEYWORDS

Peer review; reviewer network; editors; research integrity

## Introduction

Editors and peer reviewers play a crucial role as gatekeepers of the scientific literature (Tennant and Ross-Hellauer 2020) but in recent years, concerns have been raised about problematic peer review that falls short of the expected standards specified by Starck (2017). The dramatic expansion of the academic literature in the past decade (Hanson et al. 2024) has made it harder for editors to find peer reviewers, potentially leading to “reviewer fatigue” (Fox, Albert, and Vines 2017). In addition, methods introduced to incentivize peer reviewing, such as tokens toward open access charges, or public recognition on Publons, may have the unintended consequence of encouraging individuals to volunteer to review in bad faith, producing brief, superficial, plagiarized, or AI-generated reviews (Geldsetzer et al. 2023; Logullo et al. 2025; Piniewski et al. 2024).

Our focus here is on a recently discovered kind of peer review misconduct, the review mill, several cases of which were described by Oviedo-García (2024). A review mill is identified when one or more researchers produce peer reviews that appear to be derived from a common template which can be tracked across different articles and that typically include coercive citation requests that boost citations to the work of the reviewer or to others in the network. In effect, they combine two problematic practices. The first is formulaic peer review (Labbé et al. 2025), with generic, or boilerplate, comments rather than specific critique of an article; the second is coercion to cite references to work authored by the peer reviewer (Barnett 2025). When these two practices have been noted, they appear to have involved individual bad actors (Arredondo

**CONTACT** Dorothy V. M. Bishop  [dorothy.bishop@psy.ox.ac.uk](mailto:dorothy.bishop@psy.ox.ac.uk)  Department of Experimental Psychology, Life and Mind Building, South Parks Road, University of Oxford, OX1 3EL, UK

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Montero 2025; Van Noorden 2020). More co-ordinated abuses of peer review have been observed in the form of “citation rings,” defined by Lockwood (2020) as “a cabal of authors who agree to cite each other’s papers even where they are completely inappropriate.” Review mills may also involve a network of individuals working together to combine formulaic reviews with coercive citation. Just as with a citation ring, the goal appears to be to increase citations of those in the ring, but in the case of a review mill, this strategy is taken to a new level; formulaic, superficial reviews are generated, presumably to allow the participants to complete large numbers of peer reviews with minimal effort. The existence of a review mill network raises questions about how members of the network are able to co-ordinate their reviews, and whether editorial integrity has been compromised.

Review mills are typically not identifiable by readers unless peer review reports are made public, so that commonalities between review reports can be identified. Even so, they are not easy to track down, because they require in-depth scrutiny of the review reports.

Review mills have only recently been described and there is no standard approach to analyzing them. This report is an exploratory investigation documenting the methods we used to study a putative review mill in the field of gynecological oncology, explaining how we established similarity between reports that showed beyond reasonable doubt that peer reviewers were working together. Prior work has been descriptive; here we attempt to quantify similarities between reviews and to display patterns of citations across reviewers. Our goal was to develop an analytic approach that would provide insights into the characteristics that might be quantified and treated as red flags for a review mill and could potentially be used for automated detection in the future.

Behaviors characteristic of a review mill need to be clearly delineated from normal peer review practices. To that end, we did a parallel analysis of a comparative set of reviews by other reviewers of a subset of the same articles. This allowed us to ask three specific questions. First, can boilerplate comments be regarded as a common finding, perhaps based on language from instructions to reviewers? Second, is the number of non-boilerplate comments unusually low in review reports that contain boilerplate? Third, are suggested citations to the reviewer’s work unusually common in review reports with boilerplate text?

## Methods

All information used in our analysis is in the public domain. A chance observation unearthed the first review, which used what we term “boilerplate” terminology: sentences or phrases that are generic, insofar as they do not refer to content that is specific to the article, and which are found across reviews of different articles. Examples of boilerplate comments are:

- *“In my honest opinion, the topic is interesting enough to attract the readers’ attention.”*
- *“The authors have not adequately highlighted the strengths and limitations of their study. I suggest better specifying these points.”*

Subsequent review reports were identified by a snowball search, where we extended the collection of search terms as we found additional instances of boilerplate terms in every review report. We used three types of Boolean Google (search operator) searches: “reviewer” AND

- (a) A sentence or word chain from a report that met our criteria, i.e., what we term “boilerplate” wording AND/OR
- (b) PMID numbers from a report that met our criteria (see Table 1) AND/OR
- (c) Reviewers’ names from a report that met our criteria

Search results found between 7 November 2024 and 1 September 2025 were retained in a target set of possible “review mill reports” when the report contained at least one (almost) verbatim sentence. This target set was archived using the Wayback Machine (<http://wayback.archive.org/>).

We also performed an identity check on reviewers, using links from open peer reviews to a valid ORCID or the reviewer’s full name and checked their review history and H-index from Web of Science between 30 September 2025 and 7 October 2025.

**Table 1.** Numbers of articles included in the target set, by publisher and journal\*.

Publisher/journal	N articles
MDPI	
Cancers	16
Diagnostics	15
International Journal of Molecular Sciences	14
Current Oncology	5
3;Biomedicines	5
Other	20
PLOS ONE	36
BMC	
BMC Women's Health	12
BMC: Other	4
AME Publishing Company	
Translational Cancer Research	9
Gynecology and Pelvic Medicine	3
AME: Other	3
BMJ Open	11
Qeios	7
F1000 Research	4
Baishideng Publishing Group: World Journal of Clinical Cases	3
Sage	2

\*Individual journals shown for cases with three or more articles in the target set.

The whole set of boilerplate terms was scrutinized to create a coding framework, described more fully below. This was developed iteratively by identifying common sentences and word chains and grouping together those with closely similar wording. The three authors resolved by discussion how to handle cases where comments were closely similar but not identical. A separate code was used for reviewer comments that did not match our definitions of boilerplate.

Each target set report was given an ID code, and all reviewer comments were then coded using the same framework. In addition, the PMIDs or other requested citations (DOIs or titles) were listed for each review.

We cross-tabulated the occurrence of boilerplate terms against the reviewer name to create a table showing the number of instances of that term being used. A similar analysis was done to cross-tabulate PMIDs mentioned in the review report against the name of the reviewer. Reviewers are identified by numbers in this report.

A comparison group of 59 review reports was selected by focusing on the nine individuals who produced the most review reports in the target set and selecting 50% of articles they had reviewed to identify a comparison review report meeting these criteria:

- (a) consisting of at least three sentences;
- (b) the reviewer was not included in the target set; and
- (c) if possible the review was by a named reviewer (though an anonymous review was used if no other review met criteria a and b).

Thus, the comparison set of review reports was matched to the target set insofar as they involved a review report of a subset of the same articles. Any suggested citations in this comparison set of reviews were checked to see if they involved articles authored by the reviewer.

## Results

The target set contained 195 review reports, which shared verbatim or highly similar boilerplate text with other publicly available review reports. These reports belonged to 170 articles published between 6 February 2019 and 7 July 2025. This target set included 19 articles with two boilerplate reports, and three articles with three boilerplate reports. All but one of these cases involved different reviewers reviewing the same article. In the other case, our search identified three reviews by the same reviewer of the same article, corresponding to three rounds of review; this reviewer barely modified their review report other than to suggest new references to be cited. The reviewer was anonymous for 76 of the 195 reports (39%); named

reviews were contributed by 18 different reviewers. Details of articles with target set reports are available online as Supplementary Data 1.

Of the 195 boilerplate reviews, 186 suggested at least one article for citation coauthored by the reviewer or another potential review mill member. In most cases, the request for citation did not identify the authors of the recommended articles, but rather referred to them by PubMed ID number (PMID), Digital Object Identifier (DOI), or title; only 21 review reports gave any information about authors when suggesting a citation. The median number of requested citations per review was two (100 reviews), with a range from 1 to 12.

The full set of coded peer review comments is available online (Supplementary Data 2).

### Journals/publishers

Articles featuring in the target set came from several publishers and platforms. These were MDPI, PLOS, BioMed Central, AME Publishing Company, BMJ Publishing Group, Qeios, and F1000 Research (see Table 1). Note that we can detect boilerplate reviews only for journals that adopt open peer review practices.

### Cross-tabulation of reviewers with boilerplate comments

Table 2 shows the boilerplate codes and numbers of occurrences of different wordings. To simplify data presentation, codes are grouped into 13 categories.

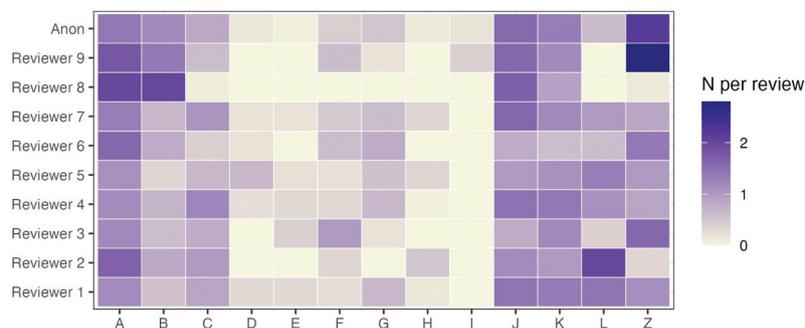
Figure 1 is a heat map where darkness of the color shows the mean number of instances of the review codes for each of nine reviewers and for anonymous reviews. The underlying data are available online (Supplementary Data 3).

### Cross-tabulation of reviewers with suggested PMIDs

We use the term “coercive citation” for comments where a potential review mill member suggests that the article should be revised to contain mention of an article by themselves or another reviewer in the target set. There were 151 individual articles suggested as coercive citations, with many articles suggested several times, including by different reviewers. The heat map in Figure 2 shows the number of coercive citations, and the superimposed asterisks show whether or not the reviewer was an author of the coerced article. Underlying data are available online (Supplementary Data 4).

We noted that when more than one reviewer from the target set was involved in reviewing an article, they typically suggested a complementary set of references for citation (see Supplementary Data 1 for examples).

We also assessed for each PMID, how often a reviewer’s suggestion for a citation was successful (see Supplementary Data 4). For 14 review reports, this could not be evaluated, because there was no indication that the article had been revised after peer review: these were articles on platforms that adopted a “preprint,



**Figure 1.** Heatmap where darkness of color indicates average number of comments of that type per review by each of nine reviewers, and by anonymous reviewers (Anon). For key to codes see Table 2. Note that the more common boilerplate codes were used by all the reviewers. The final column (Z) shows non-boilerplate comments.

**Table 2.** Coding scheme for boilerplate language in reviews. Italicized text gives some explanation. Optional phrases in square brackets; alternate wordings in pointed brackets separated by slash.

Type	Description	textual variants	N*
A	Preamble	<I read with great interest the manuscript [titled < xxx>], which>/<In my [honest] opinion, [the analyzed] topic [analyzed]> [is] interesting enough to attract the reader's attention [Dear Authors,]<I read [with great interest] the Manuscript [titled < xxx>] which falls within the aim of the Journal [and offers a highquality overview of the topic] >; (NB "highquality" as one word)	147 130
B	Prelude	I was [particularly] pleased to <revise/review> <this paper/the manuscript [en]titled < xxx> Although the manuscript can be considered already of <good>/<high>quality, I would suggest [to take into account] [the] following [minor] recommendations: [Nevertheless,] authors should clarify some point[s] [and improve the <introduction/discussion/ quality of [the] manuscript> [by] [citing relevant and novel key articles about the topic] [as suggested below] In general, the Manuscript may benefit from some < minor/major revisions>, as suggested below Authors should consider the following [minor] recommendations	6 57 84 4 64
C	Typos/formatting/English	<I suggest [another] [round of] language revision, in order to correct [a] [few] typos and improve readability [within the manuscript]>/<All the text needs a minor language revision in order to improve some typos and grammatical errors> [Moreover] <The whole text should be < corrected/revised > by a native English speaker in order to make the work clearer and more readable>/<an extensive correction/revision of the English language is necessary in order to make the work clearer and more readable> [First] Typos errors < should/must > be corrected [With no indication of what these are] [A] moderate < review/correction > of English [grammar and] language should be performed Minor editing [of English language] is <necessary/required> Manuscript should be further revised by a native English speaker [<in order to correct several typos>/ <to improve [its clarity and] readability>] Manuscript should be further revised in order to correct some typos and improve style	80 17 6 7 7 25 12
D	Abstract	I think that the abstract of this article is <very clear and well structured>/<is well organized [and clear]> The abstract of this article is not < very>clear [<and a bit too long>]/<and well structured> <So I think that > the abstract should be < completely rewritten>/<revised>in order to better summarize the contents of the manuscript The abstract < satisfactorily/perfectly > summarizes the contents of the manuscript	20 3 2 8
E	Introduction	The introduction is satisfactory [and the conclusions are supported by the data analysis] The introduction should be < extended/summarised > and completed	4 21
F	Methods	[Although it is a retrospective analysis], <criteria for inclusion and exclusion in the study>/ <inclusion/exclusion criteria > should be < better>clarified [by extending their description] [The] Methodology is accurate [</and>] <conclusions are supported by the [<data analysis/ literature review>] >/<the data analysis supports conclusions>	30 47
G	Discussion	I also suggest authors to better organize the discussion section following this ideal structure: main findings of the study, strength and limitations of the study, implications and comparison with literature, future directions <The authors have not adequately highlighted the strengths [and limitations] of their study. [I suggest better specifying these points]>/<What are the strengths and limitations of this manuscript?> <What is already known on this subject?>/<What do the results of this study add? > In my opinion, the discussion could be studied in <depth>/<deep> [and extended] <What are the actual clinical implications of this study [compared to previous ones]? <It is important to report the results obtained by the authors in the context of clinical practice and to adequately highlight what contribution this study adds to the literature already existing on the topic and to future study perspectives>/<What are the implications of these findings for clinical practice and/or further research?>	6 37 6 25 18
H	Tables/figs	<The <tables and figures>/<figures and tables > are [very] clear [and interesting]>/<Figures are interesting and tables are clear>	26
I	Ethics etc	Does this manuscript conform the Enhancing the QUALity and Transparency Of health Research (EQUATOR) network guidelines? It would be mandatory to declare about this element	12
I		Was this study registered? I could not find any information about this point	13
J	Formulaic prelude to citation request	Discussions can be expanded and improved by citing relevant articles ( <i>generic statement preceding request for PMIDs - common examples shown here</i> ). [In particular], <I suggest to>/<Authors could/should> [add further details to] discuss/consider/ <evaluate and cite> [at least briefly] {topic of PMID} [In this regard]/[Moreover] I would < recommend [to]>/<invite authors to discuss>/<discussing/ add/stress/highlight and discuss> <further details to discuss/novel evidences/novel pieces of evidence about/a reference> {topic of PMID} In my opinion author's/you have to improve the paper refering ( <i>sic</i> ) [in the text] {topic of PMID} [Maybe] it could be useful [<the evaluation of/to evaluate/to describe>] [as a comparison,]{topic of PMID} <I find [it]>/<It would be > interesting to include {topic of PMID} <i>other wordings</i>	21 22 19 8 23 41 156
K	Citation request	<i>This code is used for comments that include specific recommendation of articles - typically by citing PMIDs. The PMID can be introduced by other nonspecific statements. However, any statement that has topical content, which precedes the PMID request is coded separately (see L below).</i>	237

(Continued)

**Table 2.** (Continued).

Type	Description	textual variants	N*
L	Recommendation	<For these reasons, I recommend the publication of the article, pending few minor revisions. >/<<Considered>/<Considering>all these points, I think it could be of interest for the readers and, in my opinion, it deserves the priority [to be published after [<minor>/<major>] revisions]>/<Taking all factors into account, I <think>/<believe>it <could>/<will>be of interest to the readers and, in my <opinion>/<view>, <should be prioritized for publication>/<it deserves the priority to be published> <after>/<following>/<pending> [few] <major>/<minor>revisions>	89
Z	Not boilerplate	Because of these reasons, the article should be revised and completed <i>Use this code for any comment which does not correspond to any of the boilerplate text above.</i>	58 273

\*N is the number of occurrences of that specific boilerplate comment in the 195 target review reports.

review, publish" model, i.e., Qeios and F1000. Of the remaining review reports ( $n = 181$ ), the final article was scrutinized to check if any of the suggested citations had been incorporated in the references. One hundred and forty-two (78%) review reports led to inclusion of all or some of the suggested references in the article; in the remaining 39 (21%) authors either challenged the suggestion (e.g., by stating the reference was not relevant) or ignored it and did not incorporate the reference.

### Comparison reviews

The comparison and target set reports were similar in number of coded statements; comparison set mean = 10.7 statements ( $SD = 5.81$ ), target set mean = 9.8 statements ( $SD = 3.86$ ), a nonsignificant difference on t-test;  $t(74.1) = 1.10$ ,  $p = .274$ .

The first question addressed by analysis of comparison reviews was whether boilerplate comments may arise from reviewers following journal instructions to reviewers. If so, we should expect to see them in comparison reviews. However, there were only 16 cases of boilerplate in 633 coded comments in the comparison reviews (2%), compared with 1604 boilerplate coded comments out of 1875 coded comments in the target set reports (86%). So boilerplate comments are not a common feature of peer review reports. In fact, five boilerplate codes in the comparison set came from a single review report. A check of that reviewer's publications indicated they had coauthored articles with reviewers from the target set, suggesting that they may have been part of the same network.

Second, we wanted to assess whether the number of non-boilerplate comments was unusually low in target set reports. This was generally the case: the mean number of non-boilerplate comments for the 59 comparison review reports was 10.4 ( $SD = 5.79$ ), whereas for the 195 target set reports it was 1.4 ( $SD = 3.15$ ) ( $p < .001$  on Wilcoxon test).

Our third question was whether suggested citations to the reviewer's work are unusually frequent in the target set reports. After all, there are circumstances when it is reasonable to suggest that authors cite the reviewer's work. There were only four instances of this in the 59 reviews in the comparison set (average of 0.07 per review) compared with 370 instances in the 186 reviews in the target set (average of 1.99 per review). Suggested citations to articles other than those authored by the reviewer were rare in both groups; there were 10 review reports including such requests in the comparison set, compared to 14 in the target set. The latter typically co-occurred with suggestions to cite the reviewer's work.

### Academic editors of articles in the target set

In Supplementary Data 1, there is a spreadsheet with a row for each review identified as part of the target set. This gives details of the peer reviewer and, where it was provided by the journal, the editor. Identity of academic editors was listed for 92 out of 170 articles. In twenty-one cases, the academic editor featured in the target set (one case Reviewer 1, three cases Reviewer 5, five cases Reviewer 7, two cases Reviewer 8, and ten cases Reviewer 9).

## Characteristics of reviewers in the target set

All reviewers featuring in the target set were medical doctors with hospital and academic affiliations in Italy. They varied in productivity, as shown in Table 3. Information in this Table was largely compiled from Web of Science, supplemented by other sources such as ORCID and SciProfiles.

## Discussion

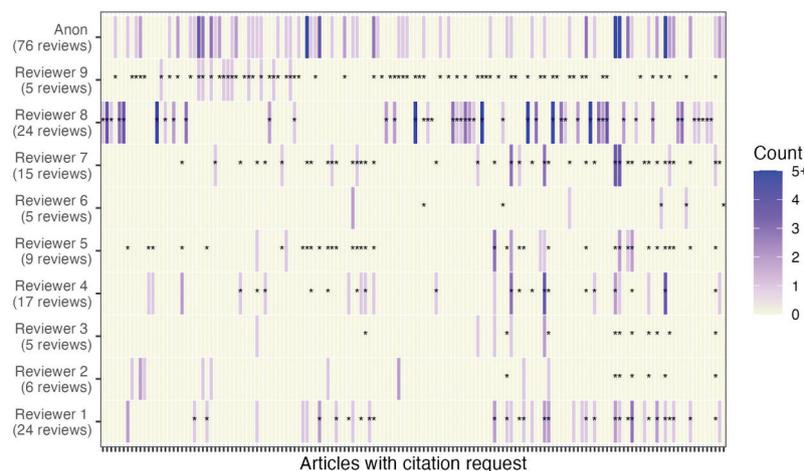
We identified 195 review reports which consisted of boilerplate language, together with requests for citations to work authored by the review millers. As shown in Figure 2, the same citations were requested across multiple reviews by different individuals.

One might wonder whether a high rate of boilerplate comments is surprising: we selected review reports on the basis that they contained boilerplate comments, so we would expect to see such phrases. The key point to note, however, is that if a review was found by searching for one boilerplate comment, then it invariably contained many others. Furthermore, the same comments recurred not just within the corpus of review reports for one reviewer, but across reports from different reviewers, who were themselves coauthors of the other reviewers. This is evident by scanning the list of comments in online Supplementary Data 2. Indeed, if one sorts this list alphabetically by comment, large stretches of comments with identical wording are readily apparent. This indicates a level of co-ordination between reviewers.

The analysis of the comparison reviews established that the review mill reports were characterized not only by presence of boilerplate comments but also by a low level of non-boilerplate comments. For many of these reviews, the only cases where a comment went beyond generic recommendations were to provide context for a suggested citation to the reviewer's work.

Day (2022) has suggested that "there are innocent reasons why referees might share template reports" and that presence of copied and pasted material across reviews does not necessarily indicate misconduct. However, an author deserves to have a peer review that engages with the content of the article; a few sentences that are duplicated across peer review reports may be acceptable but review reports that contain little or no original material engaging with the contents of a manuscript are not.

In sum, we argue that we have demonstrated beyond reasonable doubt the articles in our target set did not have acceptable peer review before they were incorporated in the corpus of academic literature. The network we identified corresponds to Oviedo-Garcia's (2024) description of a review mill, and shows that this practice extends to clinical medicine. Review mills differ from other types of fake peer review, where the peer reviewer may be an illegitimate individual who operates under a false identity (Enserink 2024). Whereas fake



**Figure 2.** Heatmap where the x-axis shows a line for each of the 151 articles that reviewers suggested, and the depth of color indicates the number of reviews where that article was suggested. Asterisks denote articles authored by the reviewer. The table containing full data used to create the heatmap, including article PMIDs, is available online as Supplementary data 4.

peer review is often found as part of a paper mill (Abalkina et al. 2025), review mills do not, as far as we are aware, involve the authors of the affected article.

### Limitations of the study

We studied a single corpus of linked review reports with the characteristics of a review mill, but it is unlikely we have found all articles related to this network, given that there were over 3400 verified peer reviews by those we identified as potential review millers registered with Web of Science up until 7 October 2025 (see Table 3). This implies that we may have identified less than 10% of the peer review reports produced by this group of researchers. Furthermore, 39% of the reports in our target set were anonymous, so there could well be other members of the network that we cannot identify. Also, without access to editorial records and open peer reviews, it is not possible to estimate what proportion of academic articles are affected by review mills; we hope that our results may inspire a fuller analysis by publishers who have access to such information.

**Table 3.** Background information about reviewers in the target set. All information was obtained from Web of Science, ORCID, or SciProfiles.

ID	H-index	Awards	Editorial Board Memberships	Verified Peer Reviews*
Reviewer 1	36	Top reviewers in Clinical Medicine - September 2018 Top reviewers for Medicine - September 2017	Minerva Obstetrics and Gynecology (current); Guest Editor of 1 MDPI Special Issue	253
Reviewer 2	15	none	Guest editor of 21 MDPI Special Issue	0
Reviewer 3	16	none	Frontiers in Medicine (current); BMC Women's Health (past membership); Heliyon (current); Journal of Clinical Obstetrics & Gynecology (current); Technology in Cancer Research & Treatment (current); Open Journal of Obstetrics and Gynecology (current) Guest editor of 4 MDPI Special Issues	99
Reviewer 4	28	none	Open Journal of Obstetrics and Gynecology (current); Guest editor of 2 MDPI Special Issues	53
Reviewer 5	42	none	Guest editor of 54 MDPI Special Issues	0
Reviewer 6	15	none	BMC Women's Health (current)	0
Reviewer 7	37	none	BMC Women's Health (current); International Journal of Gynecology & Obstetrics (current); Journal of Biological Regulators and Homeostatic Agents (current); Journal of Ovarian Research (current); PLOS ONE (current); Guest editor of 4 MDPI Special Issues	99
Reviewer 8	34	none	Reproductive Medicine	15
Reviewer 9	58	Top Peer Reviewer Top reviewers in Biology and Biochemistry - September 2019; Top reviewers in Clinical Medicine - September 2019; Top reviewers in Cross-Field - September 2019; Top reviewers in Molecular Biology and Genetics - September 2019; Academy mentor; Excellent Reviewer (51)	Diagnostics (current); International Journal of Endocrinology (current); Journal of Clinical Medicine (current); Reproductive Medicine; Acta Chirurgica Belgica (current); Acta Endocrinologica (current); BMC Women's Health (current); Gynecologic and Obstetric Investigation (current); International Journal of Fertility & Sterility (current); Italian Journal of Gynaecology and Obstetrics (current); Journal of Minimally Invasive Gynecology (current); Journal of Ovarian Research (current); Minerva Surgery (current); Minimally Invasive Therapy and Allied Technologies (current); Neuroendocrinology (current); Open Medicine (current); PLOS ONE (current); Scientific Reports (current); Translational Cancer Research (current); Wideochirurgia I Inne Techniki Maloinwazyjne (current); International Journal of Molecular Sciences (past); Experimental and Therapeutic Medicine (past); Guest editor of 20 MDPI Special Issues	2,901

\*Peer review verified through Web of Science. It is possible that a scientist will have no 'verified peer reviews' even though they engage in peer review, if they have not performed the verification step through Web of Science.

Another limitation is that we used manual methods to identify and code boilerplate comments in peer reviews. In future, computational linguistic methods to identify textual overlap of reviews would be more efficient and powerful than the manual curation and coding that we used here. Computational methods could also make it easier to map networks of reviewers and editors.

### ***Causes and consequences of review mills***

The question arises as to what motivates review millers. In the case of the network in our target set, we suspect that a goal is to increase citations and boost the reviewer's H-index, which many institutions regard as one indicator of prestige and which may lead to career advancement. We have shown that 78% of reviews achieved author compliance with all or some of the requests for citations, leading individual reviewers to gain between 4 and 57 additional citations to their work. It is noteworthy that the reviewer network that we describe is based in Italy, a country where since 2011 government funding for universities as well as recruitment and promotion of professors is determined largely by bibliometric indicators based on publications and citations (Baccini and Petrovich 2023).

Another benefit for review millers would be to obtain discounts on Article Processing Charges (APC) for future works submitted to the same publisher: for instance, MDPI offers discount vouchers to selected reviewers to compensate them for their time and effort.

Our findings suggest there are vulnerabilities in systems for reviewer selection and oversight, but we can only speculate as to how exactly review mills operate. We found that some members of the network had editorial roles. This could provide an opportunity for the editor to select reviewers from the network, which could act to increase citation of the editors' articles in a way that would be undetectable if boilerplate text had not been used. It is also noteworthy, however, that, as shown in Table 1, some of the reviews came from journals adopting the publish, review, curate model. In this model, there is no academic editor, and reviewers are selected by authors – while requiring them to avoid those that have collaborated with. These journals use automatically generated lists of potential reviewers for authors to choose from. Given that it is increasingly hard to find peer reviewers, it is likely that individuals who have a strong history of reviewing in a field, and who are known to agree readily to perform a review, will become prominent in such lists, and so gain opportunities to boost citations to of themselves or colleagues by quickly writing a boilerplate review.

Yet another way that a member of a review mill may get selected as a reviewer would be if an author suggested them. However, many journals do not accept such suggestions, and editorial scrutiny should prevent suggested reviewers being selected if they have collaborative links with the author. Overall, we did not specifically analyze relationships between authors and reviewers of articles in our target sample, but our impression was that authors were not involved in organized peer-review manipulation.

Compared to some other forms of academic malpractice, operating a review mill may seem a relatively minor peccadillo. Nevertheless, we argue that the practice should be taken seriously because of the impact review milling has on individuals and on the trustworthiness of science in general. In short, articles that have only received generic, boilerplate review comments have not undergone the kind of scrutiny that is expected from peer reviewers (Starck 2017), making the authors of the affected papers victims of research misconduct.

There is no reason to believe that original manuscripts that are reviewed by a review mill are flawed, but they are compromised by inauthentic peer review by reviewers who appear to use their position purely to suggest citations to their own work. In the case of another review mill, the publisher MDPI noted the problem with inadequately reviewed articles and proposed to re-review these (MDPI 2024); while this might help put the record straight, it can be devastating for authors to find the validity of their work questioned years after it was originally published. Indeed, a check on the status of articles in our target set on 16 February 2026 showed that two had been retracted for compromised peer review, and three had an expression of concern for “concerns about the peer review process.” All five of these articles were published in PLOS ONE, whose publisher has recently taken a proactive stance in tackling compromised peer review (Kiermer 2025).

The publisher also suffers: having been proactive in promoting transparent peer review, they then are confronted with a host of questions on how to handle papers that were published without having met their peer review criteria. Will they be able to find new peer reviewers, and if so, do they inform them that the paper was previously published? What happens if a new peer reviewer recommends rejection or major revisions? Should the Article Processing Charge (APC) be refunded, and is the agreement between publisher

and author still legally binding? Should there be a new status for the paper targeted by a review mill? What should happen to all the papers that have cited the affected work?

There is also a cost to other researchers. Academic research is a highly competitive field, and if some individuals obtain an unfair advantage by manipulating the peer review process, then other deserving individuals will lose out in terms of promotions and grants.

The most serious consequences are for the scientific endeavor: the quality of published articles will be affected when researchers cynically treat peer review as an opportunity to boost their own citations, and offer only generic comments to authors. This is particularly worrying in clinical fields, where inadequately reviewed articles and unfairly boosted articles by reviewers may pollute meta-analyses and clinical protocols, as well as leading to potentially negative impacts on future research.

### **Future directions: Tackling review mills**

In a similar way to false positive reviews, review mills are “a nightmare because they are difficult to recognize, require intensive investigations by the editor, and are suitable to corrupt the entire scientific publishing system” (Starck 2017, 34). We were able to document the phenomenon only because some publishers make peer review openly available. It would be dangerous to assume that review mills are confined to the journals we documented here: it is likely that they have infiltrated other publishers where they remain invisible and undetected. It could be argued that diligent editors should have been made suspicious by several indicators of manipulation of the peer review process as listed by the Committee on Publication Ethics (COPE) (2025), including vague and generic reviews, requests to add self-citation, and reviewers who agree to review many manuscripts. However, if editors are involved in the review mill, then it may go undetected for years. We propose that to tackle this problem we need an open peer review process, ideally with identity of reviewers and editors being available in meta-data for each article, together with the full text of peer reviews. Furthermore, if a review mill is detected, it is likely that a wide body of work is implicated, and other articles reviewed by the review millers should be checked.

We note that while publishers may be criticized for not defending better against review mills, the boilerplate reviews that we identified were not obviously problematic. When one sees large numbers of identical reviews, the problem is obvious (see Supplementary Data 2), but typically an editor would not be aware of anything amiss, especially since the identity of authors of the suggested citations is not evident unless the reference is checked. We were able to detect the boilerplate reviews from publishers listed in Table 1 because these publishers followed best practice in using open peer review. It would be unfortunate if other publishers failed to adopt open practices for fear of being found to be affected by review mills. Of course, publishers have access to all peer reviews and indeed we are aware of growing interest by publishers in running automated plagiarism checks on reviews. However, if each publisher acts independently, there remains scope for review mills to go undetected if peer review is kept hidden. There is a strong case for the STM integrity hub (International Association of Scientific, Technical & Medical Publishers 2024) to facilitate sharing peer review reports between publishers, to help combat peer review mills.

### **Conclusions**

In conclusion, peer review milling threatens the scientific record and gives an unfair career advantage to members of the review mill. Whereas most cases of compromised publication integrity focus on authors (Bolland, Avenell, and Grey 2024), in the case of review mills, authors as well as publishers are victims. In medical contexts, patient safety can be affected when clinical articles are inadequately scrutinized during peer review. Since review mill products can only be identified in those rare journals with open peer review, it is likely that this practice is far more common than suggested by our analysis. We recommend that publishers adopt open peer review, employ automated checks, and transparently report the editors responsible for handling papers, to make detection of review mills easier for the entire scientific community.

### **Disclosure statement**

No potential conflict of interest was reported by the author(s).

## Funding

The author(s) reported there is no funding associated with the work featured in this article.

## ORCID

M. Ángeles Oviedo-García  <http://orcid.org/0000-0001-5177-4899>

René Aquarius  <http://orcid.org/0000-0002-0968-6884>

Dorothy V. M. Bishop  <http://orcid.org/0000-0002-2448-4033>

## Authors' contributions

CRedit: **MAO**: Conceptualization, Data Curation, Methodology, Analysis, Writing; **RA**: Methodology, Analysis, Writing; **DB**: Data curation, Methodology, Visualization, Analysis, Writing.

## Data availability statement

The four supplementary data tables referred to in the text are available here: <https://osf.io/k5acw/overview>.

## References

- Abalkina, A., R. Aquarius, E. Bik, D. Bimler, D. Bishop, J. Byrne, G. Cabanac, A. Day, C. Labbé, and N. Wise. 2025. "Stamp Out Paper Mills"—Science Sleuths on How to Fight Fake Research." *Nature* 637 (8048): 1047–1050. <https://doi.org/10.1038/d41586-025-00212-1>.
- Arredondo Montero, J. 2025. "Rethinking Peer Review: Ethics and Blind Spots of Scientific Gatekeeping." *World Journal of Pediatrics* 21 (9): 872–874. <https://doi.org/10.1007/s12519-025-00960-0>.
- Baccini, A., and E. Petrovich. 2023. "A Global Exploratory Comparison of Country Self-Citations 1996–2019." *PLOS ONE* 18 (12): e0294669. <https://doi.org/10.1371/journal.pone.0294669>.
- Barnett, A. 2025. "Are Peer Reviewers Influenced by Their Work Being Cited?" *eLife* 14, RP108748. <https://doi.org/10.7554/eLife.108748>.
- Bolland, M. J., A. Avenell, and A. Grey. 2024. "Publication Integrity: What Is It, Why Does It Matter, How It Is Safeguarded and How Could We Do Better?" *Journal of the Royal Society of New Zealand* 1–20. <https://doi.org/10.1080/03036758.2024.2325004>.
- COPE. 2025. How to Recognise Potential Manipulation of the Peer Review Process, Committee on Publication Ethics. April 7. <https://publicationethics.org/guidance/flowchart/how-recognise-potential-manipulation-peer-review-process>.
- Day, A. 2022. "Exploratory Analysis of Text Duplication in Peer-Review Reveals Peer-Review Fraud and Paper Mills." *Scientometrics* 127 (10): 5965–5987. <https://doi.org/10.1007/s11192-022-04504-5>.
- Enserink, M. 2024. 'It Felt Very Icky': This Scientist's Name Was Used to Write Fake Peer Reviews. December 3. <https://www.science.org/content/article/it-felt-very-icky-scientist-s-name-was-used-write-fake-peer-reviews>.
- Fox, C. W., A. Y. K. Albert, and T. H. Vines. 2017. "Recruitment of Reviewers Is Becoming Harder at Some Journals: A Test of the Influence of Reviewer Fatigue at Six Journals in Ecology and Evolution." *Research Integrity and Peer Review* 2 (1): 3. <https://doi.org/10.1186/s41073-017-0027-x>.
- Geldsetzer, P., M. Heemann, P. Tikka, G. Wang, M. M. Cusick, A. Lenjani, and N. Krishnan. 2023. "Prevalence of Short Peer Reviews in 3 Leading General Medical Journals." *JAMA Network Open* 6 (12): e2347607. <https://doi.org/10.1001/jamanetworkopen.2023.47607>.
- Hanson, M. A., P. G. Barreiro, P. Crosetto, and D. Brockington. 2024. "The Strain on Scientific Publishing." *Quantitative Science Studies* 5 (4): 823–843. [https://doi.org/10.1162/qss\\_a\\_00327](https://doi.org/10.1162/qss_a_00327).
- International Association of Scientific, Technical & Medical Publishers. 2024. *STM Integrity Hub*. STM Association. <https://stm-assoc.org/what-we-do/strategic-areas/research-integrity/integrity-hub/>.
- Kiermer, V. 2025. "PLOS responds to PNAS study detailing the growth of peer review integrity issues." The Official PLOS Blog. August 11. <https://theplosblog.plos.org/2025/08/plos-responds-to-pnas-study-detailing-the-growth-of-peer-review-integrity-issues/>.
- Labbé, C., G. Hubert, W. Swart, and G. Cabanac. 2025. "Evidence of Use of Template-Based Peer-Review Reports and Concern About Review Mills." *PRC'25: 10th International Congress on Peer Review and Scientific Publication*: Hal-05273020. September. <https://hal.science/hal-05273020v1>.
- Lockwood, M. 2020. "Editorial: Citation Malpractice." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 476 (2242): 20200746. <https://doi.org/10.1098/rspa.2020.0746>.
- Logullo, P., A. MacCarthy, S. Kirtley, G. S. Bullock, P. Dhiman, J. Ma, and G. S. Collins. 2025. "Peer Review Reports of Randomized Controlled Trials in Oncology Can Be Short and Superficial." *Journal of Clinical Epidemiology*: 185. <https://doi.org/10.1016/j.jclinepi.2025.111893>.

- MDPI. 2024. Interim Report on Potential Misconduct by Reviewers Involving 84 Papers. February 13. <https://www.mdpi.com/about/announcements/7880>.
- Oviedo-García, M. Á. 2024. "The Review Mills, Not Just (Self-)Plagiarism in Review Reports, but a Step Further." *Scientometrics* 129 (9): 5805–5813. <https://doi.org/10.1007/s11192-024-05125-w>.
- Piniewski, M., I. Jarić, D. Koutsoyiannis, and Z. W. Kundzewicz. 2024. "Emerging Plagiarism in Peer-Review Evaluation Reports: A Tip of the Iceberg?" *Scientometrics* 129 (4): 2489–2498. <https://doi.org/10.1007/s11192-024-04960-1>.
- Starck, J. M. 2017. "Ethical Rules in the Reviewing Process." In *Scientific Peer Review: Guidelines for Informative Peer Review*, edited by J. M. Starck, 33–34. Wiesbaden, Germany: Springer Fachmedien. [https://doi.org/10.1007/978-3-658-19915-9\\_7](https://doi.org/10.1007/978-3-658-19915-9_7).
- Tennant, J. P., and T. Ross-Hellauer. 2020. "The Limitations to Our Understanding of Peer Review." *Research Integrity and Peer Review* 5 (1): 6. <https://doi.org/10.1186/s41073-020-00092-1>.
- Van Noorden, R. 2020. "Highly Cited Researcher Banned from Journal Board for Citation Abuse." *Nature* 578 (7794): 200–201. <https://doi.org/10.1038/d41586-020-00335-7>.