

# Plenary Session Abstracts

Note: Abstracts reflect the status of the research at the time the abstracts were accepted for presentation.

## Thursday, September 8, 2022

### Authorship, Contributorship, and Misconduct

#### Prevalence of Honorary Authorship According to Different Authorship Recommendations and Contributor Role Taxonomy (CRediT) Statements

Nicola Di Girolamo,<sup>1,2</sup> Reint Meursinge Reynders,<sup>3,4</sup> Vincent Lariviere,<sup>5</sup> Mostafa Ibrahim<sup>6</sup>

**Objective** The International Committee of Medical Journal Editors (ICMJE) provided a set of minimum criteria for authorship.<sup>1</sup> These recommendations have been adapted for all sciences in an article by McNutt and colleagues.<sup>2</sup> The main difference between these 2 sets of recommendations is that the science-wide recommendations do not require authors to draft or revise the work.<sup>1,2</sup> This study aimed to identify the proportion of authors who, based on their self-compiled Contributor Role Taxonomy (CRediT) statements, did not meet the minimum criteria for authorship (ie, were honorary authors) according to these 2 sets of recommendations. Furthermore, the study aimed to identify the proportion of authors who supplied only resources and/or funding to a study. Such authors were identified as “supply authors,” and this practice is considered to be a subtype of honorary authorship.

**Design** Cross-sectional study of CRediT statements published in scholarly articles. The Public Library of Science (PLOS) provided CRediT statements and associated data for authors of articles published in PLOS journals between July 2017 and October 2021. Two investigators independently evaluated the authorship recommendations<sup>1,2</sup> and developed logical operations of CRediT items for each recommendation. A third investigator acted as an arbiter in case of disagreement. The criteria related to approval and accountability for both recommendations could not be verified because of current CRediT items. For the second objective of the study, authors who contributed only to roles funding acquisition and/or resources were identified.

**Results** A total of 629,046 CRediT statements (1 per author) originating from 82,683 journal articles were included. Of the CRediT statements, 34.8% ( $n = 218,563$ ; 95% CI, 34.7%-34.9%) indicated that the contributions provided by the author were not sufficient to qualify for authorship according to the ICMJE recommendations. Based on science-wide recommendations, 3.6% ( $n = 22,575$ ; 95% CI, 3.5%-3.6%) of the authors did not qualify for authorship. Sensitivity analyses accounting for potentially ambiguous CRediT items

provided similar results. The odds of fulfilling only 1 of the recommendations steadily decreased from 2017 to 2021 (Table 1), and authors of articles published in nonmedical journals had 1.11 times the odds to fulfill only 1 of the recommendations compared with authors of articles published in medical journals. Overall, 8394 authors (1.33%; 95% CI, 1.31%-1.36%) were “supply authors.” Their prevalence decreased in the years from 2017 to 2019 (difference in proportion, 0.6%; 95% CI, 0.57%-0.63%) but remained unchanged from 2019 to 2021 (0%; 95% CI, -0.02% to 0.02%).

**Conclusions** Based on self-compiled CRediT statements, honorary authorship is still prevalent in science, although it seems to have steadily decreased in recent years. A seemingly minor edit applied to the ICMJE recommendations resulted in substantially different authorship requirements. Efforts should be directed toward developing consensus on core tasks to qualify for authorship that are widely applicable in science. Additional strategies should be implemented to address “supply authorship.”

**Table 1. Prevalence of Honorary Authorship Based on CRediT Statements Depending on 2 Authorship Recommendations and the Association of Year and Journal Area With Disagreements Between Recommendations**

	ICMJE recommendations, No./total No. (%)	Science-wide recommendations, No./total No. (%)	Adjusted OR (95% CI) <sup>a</sup>
<b>Year</b>			
2017	28,680/75,439 (38.0)	3239/75,439 (4.3)	1 [Reference]
2018	56,043/155,548 (36.0)	5836/155,548 (3.8)	0.94 (0.92-0.96)
2019	46,931/137,397 (34.2)	4703/137,397 (3.4)	0.87 (0.86-0.89)
2020	49,611/145,820 (34.0)	4905/145,820 (3.4)	0.87 (0.85-0.89)
2021	37,598/114,842 (32.7)	3892/114,842 (3.4)	0.82 (0.80-0.84)
<b>Journal area<sup>b</sup></b>			
Medical	21,481/66,310 (32.4)	2134/66,310 (3.2)	1 [Reference]
Nonmedical	197,382/562,736 (35.1)	20,441/562,736 (3.6)	1.11 (1.09-1.13)

Abbreviations: ICMJE, International Committee of Medical Journal Editors; OR, odds ratio.

<sup>a</sup>Values are provided from the multivariable model including authors with disagreement between recommendations as the dependent variable and year and journal area as independent categorical variables. For all comparisons,  $P < .001$ .

<sup>b</sup>For analytical purposes, journals that had medicine as one of their subject areas on the Scimago Journal Rank were classified as medical, whereas other journals were classified as nonmedical.

## References

1. International Committee of Medical Journal Editors (ICMJE). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated December 2021. Accessed February 27, 2022. <http://www.icmje.org/icmje-recommendations.pdf>
2. McNutt MK, Bradford M, Drazen JM, et al. Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proc Natl Acad Sci U S A.* 2018;115(11):2557-2560. doi:10.1073/pnas.1715374115

<sup>1</sup>College of Veterinary Medicine, Cornell University, Ithaca, NY, USA, nicoladiggi@gmail.com; <sup>2</sup>*Journal of Small Animal Practice*, British Small Animal Veterinary Association, Gloucestershire, UK; <sup>3</sup>Department of Oral and Maxillofacial Surgery, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, the Netherlands; <sup>4</sup>Private practice of orthodontics, Milan, Italy; <sup>5</sup>École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal, QC, Canada; <sup>6</sup>School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA

**Conflict of Interest Disclosures** Nicola Di Girolamo is an editor in chief of 2 peer-reviewed journals, 1 published by Elsevier and 1 by Wiley. No other disclosures were reported.

**Additional Information** The protocol of this research study was registered in the Open Science Framework (<https://osf.io/ezpxs/>).

## Use of an Artificial Intelligence-Based Tool for Detecting Image Duplication Prior to Manuscript Acceptance

Daniel S. Evanko<sup>1</sup>

**Objective** Image reuse is a common problem with the integrity of image data reported in scientific articles. Considerable resources are expended by the community in trying to detect, communicate, and respond to potential image reuse. These efforts are manually intensive and often take place after publication, when resolution is difficult and time-consuming. The study objective was to evaluate the use of an artificial intelligence-based tool to identify potential image reuse prior to acceptance and address problems then.

**Design** In this cross-sectional study designed to systematically identify and act on instances of image reuse in manuscripts submitted to 9 journals published by the American Association for Cancer Research, a machine-assisted process was implemented for detecting potential image duplication between images in a manuscript. Prior to issuing a provisional acceptance decision, an internal editor selected original research manuscripts containing images susceptible to duplication and detection and uploaded them to a commercial tool (Proofig). The tool's assessment was followed by editor evaluation and refinement of the results and communication of potentially problematic image duplications to the authors using a standard report generated by the tool. Author responses and the outcomes of these

**Table 2. Author Responses to the Report of Image Duplication and Outcome of the Author Query**

Author response and outcome <sup>a</sup>	No. (%) of manuscripts (N = 207)
Unintentional duplication	131 (63)
Explanation added to text	3 (1.4)
Image(s) replaced	119 (57.5)
Text and image changes	9 (4.3)
Duplicated for presentation purposes	58 (28)
Explanation added to text	44 (21.3)
Image(s) replaced	5 (2.4)
Text and image changes	9 (4.3)
Intentional inappropriate duplication	1 (0.5)
New experiment	1 (0.5)
No actual duplication	12 (5.8)
Explanation added to text	4 (1.9)
Image(s) replaced	2 (1.0)
No change made	6 (2.9)
Author investigating or no explanation provided	5 (2.4)
Manuscript rejected	1 (0.5)
Manuscript withdrawn	4 (1.9)

<sup>a</sup>A single response and outcome was recorded for each author query regardless of the number of separate duplications in the manuscript. If there were different explanations for different duplications, only the most serious response and outcome were recorded.

queries were recorded in a tracking sheet using a standardized classification system.

**Results** From January 2021 through May 2022, a total of 207 image duplication queries were sent to authors for their response. This represented 9.3% of 2220 original research manuscripts that reached this editorial stage and 15% of the 1367 manuscripts selected for analysis. The distribution of duplicate quantity per manuscript was 104 (50%) with 1 duplicate, 46 (22%) with 2 duplicates, 19 (9%) with 3 duplicates, 28 (14%) with 4 to 10 duplicates, and 11 (5%) with 11 or more duplicates. Responses from authors indicated that 63% (n = 131) of duplications were unintentional, for example, from general image mishandling, and 28% (n = 58) were intentional duplications for presentation purposes (see **Table 2** for details). In 2% (n = 5) of cases the author said they were investigating or provided no explanation. These manuscripts were withdrawn or rejected. In all other cases, changes were made to address the duplication(s). Only in 12 cases did the potential duplication turn out not to be a duplication. To compare the time required to identify seemingly real duplications for communication to authors, 5 editors first analyzed 27 manuscripts purely manually and then using the tool followed by manual evaluation and refinement. The mean time per manuscript for tool-assisted analysis was 4.4 minutes (11 duplications identified in 8 manuscripts) vs 8 minutes (5 duplications identified in 3 manuscripts) for manual analysis.

**Conclusions** The use of this artificial intelligence-based tool effectively identified real duplications between and

within figures in 14% (195/1367) of manuscripts intended for acceptance that contained images susceptible to duplication and detection. This allowed these problems to be addressed prior to publication with minimal manual effort.

<sup>1</sup>American Association for Cancer Research, Philadelphia, PA, USA, daniel.evanko@aacr.org

**Conflict of Interest Disclosures** Daniel S. Evanko is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

## Publication and Collaboration Anomalies in Academic Papers Originating From a Russian-Based Paper Mill

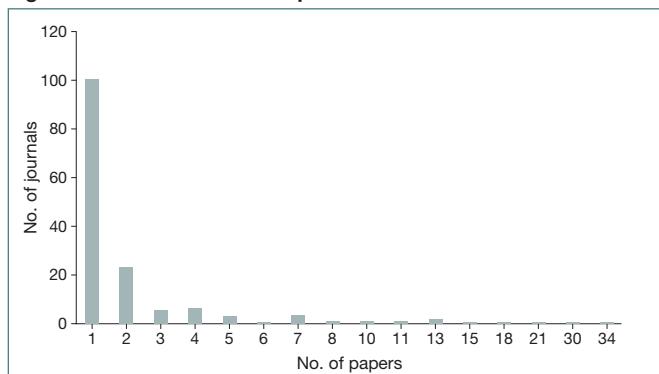
Anna Abalkina<sup>1</sup>

**Objective** Paper mills represent an offer for on-demand writing of fraudulent academic manuscripts for a fee. There is evidence of increasing infiltration of the academic literature with fraudulent papers originating from paper mills. The majority of known paper mills originate from China and less is known about their operations in other countries, namely Iran and Russia, where paper mills also operate.<sup>1</sup> This study attempted to shed light on the activity of paper mills using the experience of a Russian mill to detect papers originating from the Russian-based paper mill International Publisher LLC and to identify a set of factors associated with fraudulent papers.

**Design** Offers published during 2019–2021 were collected from the 123mi.ru website. The details of the offer as the title/topic of the paper, number of coauthors, country/region of a journal, date of publication, and in some cases country of the author and abstract were analyzed to identify auctioned papers. Many fraudulent papers were detected because they were published with identical or closely worded titles. The correctness of detection was confirmed by matches of other offer details.

**Results** A total of 1009 offers were published during the study period, and the study identified at least 436 papers (43%) published in 154 journals potentially linked to the paper mill as of mid-March 2022; 22 of the papers appeared in 3 hijacked journals. More than 800 scholars from at least 39 countries purchased coauthorships from this paper mill.

**Figure 1. Number of Articles per Individual Journal**



In addition, 22 articles were published in 3 hijacked journals.

Fraudulent papers were published in predatory journals, journals of reputable publishers, and hijacked journals that represent cybercriminal publishers who clone titles and other metadata of legitimate journals (**Figure 1**). The list of identified papers can be accessed via a shared spreadsheet.<sup>2</sup> This study found collaboration anomalies in questionable papers in which authors did not have common research interests, specialized in different disciplines, or were affiliated with different universities that did not focus research on the paper's subject or might not specialize in the topic of the paper. There was also evidence of data fabrication in these papers. The study detected dishonest collaboration with journals (purchase of an entire issue) or editors who were listed as coauthors of fraudulent papers. This paper mill also applied a “one paper–one journal” principle, ie, submission of a problematic paper to an individual legitimate journal only once.

**Conclusions** The production of this paper mill was difficult to detect due to individually tailored papers being submitted to 154 journals. Journals themselves had no opportunity to notice irregularities from 1 paper. Detection of irregularities can require regular upgrades of the system of detection of fraudulent papers by publishers and COPE (Committee on Publication Ethics), which is based on tracing similarity patterns among manuscripts. Because the study analyzed a single paper mill, it is likely that the number of papers with forged authorship is much higher.

## References

- Else H, Van Noorden R. The fight against fake-paper factories that churn out sham science. *Nature*. 2021;591:516–519. doi:10.1038/d41586-021-00733-5
- Abalkina A. Paper mill “International Publisher.” <https://docs.google.com/spreadsheets/d/1vzjtRPX7kd2KczdtKONEpRZb2F-4lj5Sd9jL6DfbiBk/edit?usp=sharing>

<sup>1</sup>Freie Universität Berlin, Berlin, Germany, anna.abalkina@fu-berlin.de

**Conflict of Interest Disclosures** None reported.

## Effect of Alerting Authors of Systematic Reviews and Guidelines That Research They Cited Had Been Retracted: A Randomized Controlled Trial

Alison Avenell,<sup>1</sup> Mark J. Bolland,<sup>2</sup> Greg D. Gamble,<sup>2</sup> Andrew Grey<sup>2</sup>

**Objective** Retracted clinical trials may be influential in systematic reviews and guidelines. Fanelli et al<sup>1</sup> examined the impact of retractions on 50 meta-analyses and found little influence but suggested that the impact was likely variable and context specific. Using retracted clinical trial reports from one research group,<sup>2</sup> responses of editors and authors to notifications that systematic reviews and guidelines they had published had cited the retracted trials were evaluated.

**Design** Between November 2019 and January 2020, for 27 retracted trials (published 1997 to 2012, retracted 2016 to 2019) in osteoporosis and neurology, searches were

conducted in Web of Science, Scopus, and researchers' files for systematic reviews and clinical guidelines that cited the trials as evidence. Citing publications that acknowledged that they had cited retracted work were excluded. For each citing publication, 2 researchers independently coded the likely impact of removing the retracted trial reports, including findings likely to change (yes, no, or uncertain) and size of change (substantial, moderate, or minor). In a factorial design with 4 groups, authors of citing publications were randomized to receive up to 3 emails (if no reply) to the contact author and journal editor vs contact author only and, for citing publications with 2 or more authors, to an email to the contact author only vs up to 3 authors (selected from among contact, first, second, and final authors). Emails giving details of the retracted trials were sent monthly September to October 2020. Follow-up was 1 year for outcomes, assessed by replies to emails and notices in the public domain, including any reply (yes or no), time for a reply from first author, time for any reply, and action taken. Comparisons were undertaken using  $\chi^2$  tests. Email replies were analyzed for content by 3 researchers.

**Results** A total of 88 citing publications (published 2003 to 2020) were identified; 2 were corrected before emails were sent. Authors/editors were emailed about 86 citing publications. A total of 45 of 88 citing publications (51%) had findings coded likely to change if the retracted trials were removed, and 39 of these (87%) were likely substantial impacts. Replies were received for 44 of 86 citing publications (51%). Emailing 3 authors was more likely to elicit a reply than emailing the contact author alone (26 of 42 [62%] vs 16 of 40 [40%];  $P = .03$ ), but including the editor did not increase replies (23 of 44 [52%] with editor vs 21 of 44 [48%] without editor;  $P = .66$ ). Including more authors and/or the editor, whether findings were coded likely to change and the size of the likely change, had no effect on published corrective action. One year after emails were sent, only 9 publications had been corrected.

**Conclusions** Retracted trials impact systematic reviews and guidelines. Emailing more authors to notify them of the retractions yielded more replies but did not increase corrections. Email alerts to authors and editors are inadequate to correct the impact of retracted publications in citing systematic reviews and guidelines. Publications with retracted citations should be marked until authors resolve concerns.

## References

1. Fanelli D, Wong J, Moher D. What difference might retractions make? an estimate of the potential epistemic cost of retractions on meta-analyses. *Account Res*. Published online July 14, 2021. doi:10.1080/08989621.2021.1947810
2. Bolland MJ, Avenell A, Gamble GD, Grey A. Systematic review and statistical analysis of the integrity of 33 randomized controlled trials. *Neurology*. Published online November 9, 2016. doi:10.1212/WNL.oooooooooooo0003387

<sup>1</sup>Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland, a.avenell@abdn.ac.uk; <sup>2</sup>Department of Medicine, University of Auckland, Auckland, New Zealand

**Conflict of Interest Disclosures** None reported.

**Funding/Support** Mark J. Bolland is the recipient of an HRC Clinical Practitioner Fellowship. The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Government Health and Social Care Directorates.

## Diversity, Equity, and Inclusion

### Women's Representation Among Peer Reviewers of Medical Journals

Ana-Catarina Pinho-Gomes,<sup>1,2</sup> Amy Vassallo,<sup>3</sup> Mark Woodward,<sup>1,3</sup> Sanne A. E. Peters<sup>1,3,4</sup>

**Objectives** To investigate women's representation among peer reviewers of medical journals overall and according to the gender of the editor in chief and women's representation in the editorial board.

**Design** This cross-sectional study included journals of the BMJ Publishing Group that reported the names of their peer reviewers in 2020. For each journal, the gender of the editor in chief, deputy editors, and associate editors was determined based on photographs and pronouns available on the journal website or professional affiliations. The package genderizer in R was used to predict the gender of the peer reviewers based on given names, which were extracted from full names and assigned as woman or man.

**Results** Overall, this study included 47 of the 74 journals in the BMJ Publishing Group because data were not publicly available for the remaining journals (**Table 3**). Women accounted for 30.2% of the 42,539 peer reviewers, with marked variation ranging from 8% in the *Journal of ISAKOS* to 50% in *Medical Humanities*. Women represented 33.4% of the 555 editors, including 19.2% of the 52 editors in chief. There were 5 journals with more than 1 editor in chief, all of which had 2 men as editors in chief. There were 5 journals with no woman among the editors and 12 journals in which women's representation as editors was 50% or greater. Among those 12 journals, 7 had a woman as editor in chief. There was a moderate positive correlation between the percentage of women as editors and as reviewers (Spearman correlation coefficient, 0.590;  $P < .001$ ). The percentage of women as editors excluding editors in chief was higher when the editor in chief was a woman than a man (53.3% vs 29.2%;  $P < .001$ ). The percentage of women as peer reviewers was also higher in journals with a woman as editor in chief compared with a man (32.0% vs 26.4%;  $P < .001$ ). There was no significant correlation between women's representation and the journal CiteScore (Spearman correlation coefficient, -0.288;  $P = .07$ ) or impact factor (Spearman correlation coefficient, -0.343;  $P = .09$ ). This study has some limitations, such as using binary prediction of gender based on given names, relying on data for accepted rather than invited

**Table 3. Representation of Women Among Peer Reviewers and Editors of Medical Journals<sup>a</sup>**

BMJ Publishing Group journals	Reviewers, No.	Women Reviewers, %	Missing, %	Editors, No.	Women Editors, %	Gender of EIC	CiteScore	Impact factor
<i>Annals of the Rheumatic Diseases</i>	529	23.1	0.4	12	25.0	Man	28.7	19.1
<i>BMJ Case Reports</i>	7179	23.1	1.1	11	27.3	Woman	NA	NA
<i>BMJ Global Health</i>	1325	41.1	0.8	16	25.0	Man	5.5	5.6
<i>BMJ Health &amp; Care Informatics</i>	133	34.1	0.8	17	35.3	Man	1.9	NA
<i>BMJ Leader</i>	162	47.8	1.9	14	35.7	Man	1	NA
<i>BMJ Neurology Open</i>	85	32.9	0.0	8	25.0	Man	NA	NA
<i>BMJ Open</i>	13,041	36.4	1.3	14	50.0	Man	3.7	2.7
<i>BMJ Open Diabetes Research &amp; Care</i>	1038	30.8	0.9	8	0.0	Man	3.3	3.4
<i>BMJ Open Ophthalmology</i>	278	30.1	0.7	29	34.5	Man	2.5	NA
<i>BMJ Open Quality</i>	42	39.0	2.4	8	87.5	Woman	1.1	NA
<i>BMJ Open Respiratory Research</i>	340	24.6	1.8	3	0.0	Men (2)	4	NA
<i>BMJ Open Science</i>	43	37.2	0.0	18	44.4	Woman	NA	NA
<i>BMJ Open Sport &amp; Exercise Medicine</i>	309	33.4	0.3	39	33.3	Man	3.5	NA
<i>BMJ Paediatrics Open</i>	356	35.0	0.6	26	46.2	Man	2.5	NA
<i>BMJ Simulation &amp; Technology Enhanced Learning</i>	180	44.4	0.0	12	58.3	Woman	1.4	NA
<i>BMJ Supportive &amp; Palliative Care</i>	417	48.3	0.7	29	34.5	Men (2)	4.8	3.6
<i>British Journal of Ophthalmology</i>	1113	24.5	0.3	3	0.0	Man	7.3	4.6
<i>British Journal of Sports Medicine</i>	693	28.5	0.1	15	40.0	Man	19.2	13.8
<i>Drug and Therapeutics Bulletin</i>	64	31.3	0.0	12	33.3	Man	NA	NA
<i>Emergency Medicine Journal</i>	767	26.5	0.0	6	50.0	Woman	3.4	2.8
<i>European Journal of Hospital Pharmacy</i>	203	40.5	1.5	16	37.5	Man	1.6	1.7
<i>Evidence-Based Medicine</i>	271	33.3	1.5	11	63.6	Man	3.2	NA
<i>Evidence-Based Mental Health</i>	64	35.9	1.6	12	25.0	Man	8.6	8.5
<i>Frontline Gastroenterology</i>	220	19.5	0.0	11	9.1	Man	3.2	NA
<i>General Psychiatry</i>	167	25.7	0.0	10	10.0	Man	4.5	NA
<i>Gut</i>	1307	20.2	0.8	17	5.9	Man	35.6	23.1
<i>Heart</i>	970	23.0	0.4	17	23.5	Woman	9	6.0
<i>Injury Prevention</i>	282	38.6	1.8	7	57.1	Woman	3.7	2.4
<i>Integrated Healthcare Journal</i>	35	37.1	0.0	2	0.0	Man	NA	NA
<i>Journal of Clinical Pathology</i>	441	30.9	1.8	10	30.0	Man	5.3	3.4
<i>Journal of Epidemiology &amp; Community Health</i>	548	40.7	1.5	22	27.3	Men (2)	6.3	3.7
<i>Journal of Investigative Medicine</i>	366	24.9	0.3	27	18.5	Man	3.9	2.9
<i>Journal of Medical Ethics</i>	726	38.7	0.4	8	62.5	Man	4	2.9
<i>Journal of Medical Genetics</i>	504	38.3	0.0	6	33.3	Man	9.7	6.3
<i>Journal of Neurointerventional Surgery</i>	788	11.2	0.5	16	12.5	Man	8.2	5.8
<i>Journal of Neurology, Neurosurgery, and Psychiatry</i>	1126	19.1	0.7	8	12.5	Man	13.5	10.3
<i>Medical Humanities</i>	198	50.5	1.0	5	60.0	Woman	1.5	NA
<i>Occupational and Environmental Medicine</i>	440	40.6	0.0	15	33.3	Man	6.8	4.4
<i>Open Heart</i>	365	19.2	0.3	13	23.1	Man	3.1	NA
<i>Postgraduate Medical Journal</i>	429	24.9	1.6	12	16.7	Man	3.3	2.4
<i>Practical Neurology</i>	118	16.2	0.8	6	0.0	Men (2)	3.1	NA
<i>Regional Anaesthesia &amp; Pain Medicine</i>	405	21.4	1.0	12	8.3	Men (2)	7.9	6.3
<i>RMD Open</i>	424	32.9	1.2	8	50.0	Man	6.1	5.1
<i>Journal of ISAKOS</i>	165	8.0	1.2	3	33.3	Man	NA	NA
<i>Tobacco Control</i>	519	40.9	1.2	8	75.0	Woman	10.9	6.6
<i>Trauma Surgery &amp; Acute Care Open</i>	140	26.1	1.4	10	50.0	Man	1.3	NA
<i>The BMJ</i>	3224	29.5	0.8	15	80.0	Woman	6.9	38.9
Overall	42,539	30.2	0.9	555	33.4	19.2%		

Abbreviations: EIC, editor in chief; NA, not available.

<sup>a</sup>All data refer to 2020 apart from data for *The BMJ*, which are from 2017, because this was the most recent year available.

reviewers, and not accounting for the number of reviews performed by each individual.

**Conclusions** Women account for less than 1 in 3 peer reviewers of medical journals. Women's representation as peer reviewers is higher in journals with a higher percentage of women as editors or with a woman as editor in chief, suggesting that increasing women's representation as editors may be one of many necessary steps toward gender equity. The gender gap among peer reviewers of medical journals mimics that previously found among authors and editors,<sup>1-3</sup> thus emphasizing the need to address the persisting gender gap at all levels of the publishing system.

## References

1. Lundine J, Bourgeault IL, Clark J, et al. The gendered system of academic publishing. *Lancet*. 2018;391(10132):1754-1756. doi:10.1016/S0140-6736(18)30950-4
2. Hart KL, Perlis RH. Trends in proportion of women as authors of medical journal articles, 2008-2018. *JAMA Intern Med*. 2019;179(9):1285-1287. doi:10.1001/jamainternmed.2019.0907
3. Pinho-Gomes AC, Vassallo A, Thompson K, et al. Representation of women among editors in chief of leading medical journals. *JAMA Netw Open*. 2021;4(9):e2123026. doi:10.1001/jamanetworkopen.2021.23026

<sup>1</sup>The George Institute for Global Health, School of Public Health, Imperial College London, London, UK, a.pinho-gomes@imperial.ac.uk; <sup>2</sup>School of Population Health & Environmental Sciences, Faculty of Life Sciences & Medicine, King's College London, London, UK; <sup>3</sup>The George Institute for Global Health, University of New South Wales, Sydney, New South Wales, Australia; <sup>4</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

**Conflict of Interest Disclosures** Ana-Catarina Pinho-Gomes reported receiving funding through an Academic Clinical Fellowship by the National Institute for Health Research. Mark Woodward is a consultant for Amgen, Kyowa Kirin, and Freeline and reported receiving grants from the National Health and Medical Research Council of Australia. Sanne A. E. Peters reported receiving support through a UK Medical Research Council Skills Development Fellowship. No other disclosures were reported.

## Association Between International Editorial Staff and International Publications in Leading Biomedical Journals

Gandolina Melhem,<sup>1</sup> Chris A. Rees,<sup>2,3</sup> Bruno F. Sunguya,<sup>4</sup> Mohsin Ali,<sup>5</sup> Anura Kurpad,<sup>6</sup> Christopher P. Duggan<sup>7,8</sup>

**Objective** To examine the association of having editorial staff members affiliated with low- and middle-income countries with publications from low- and middle-income countries in leading biomedical journals. It was hypothesized that greater representation from low- and middle-income countries among editorial staff would be associated with more published articles from low- and middle-income countries.

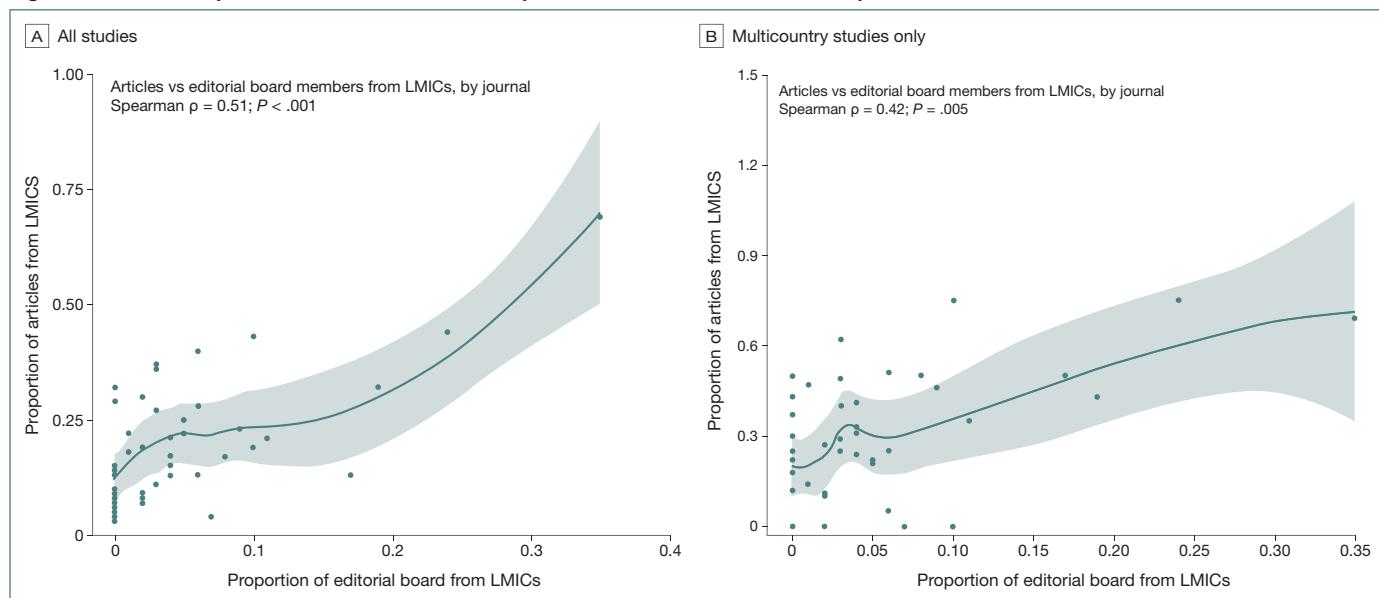
**Design** A cross-sectional study was conducted of biomedical journals in 2020 in fields whose content covers the largest disease burden globally. To obtain editorial staff country affiliations, webpages of the 5 leading journals in each of the following fields were reviewed: general medicine, pediatrics, surgery, obstetrics and gynecology, cancer, cardiovascular diseases, infectious diseases, psychiatry, and nutrition. Original research articles in each journal were reviewed through MEDLINE. The study country of each original research article (ie, where the study was conducted) was determined by searching the article title, abstract, keyword, and medical subject heading using EndNote, and by 2 authors reviewing the full text of each article to assign a study country. Editorial staff country affiliations and study country location(s) were classified according to the World Bank income brackets and regions. Descriptive statistics were used to describe the proportion of editorial staff affiliated with each income bracket and region. Spearman  $\rho$  was used to assess the relation between the proportion of editorial staff affiliated with low- and middle-income countries and published articles reporting work conducted in these countries.

**Results** There were 3819 editorial staff in the 45 included journals: 3637 (95.2%) were affiliated with high-income countries, 140 (3.7%) with upper-middle income countries, 37 (1.0%) with lower-middle income countries, and 5 (0.1%) with low-income countries. Every editor in chief ( $n = 48$ ; 100.0%) was affiliated with a high-income country. Of the 459 associate editors, 445 (96.9%) were affiliated with high-income countries, 10 (2.2%) with upper-middle income countries, 4 (0.9%) with lower-middle income countries, and 0 with low-income countries. Editorial staff were most commonly affiliated with North American (2120 [55.5%]) and European and Central Asian countries (1256 [32.9%]). Of the 10,096 original research articles included, 7857 (77.8%) reported research conducted in high-income countries, 1562 (15.5%) in upper-middle income countries, 507 (5.0%) in lower-middle income countries, and 170 (1.7%) in low-income countries. Greater editorial staff representation correlated moderately with more published articles reporting research conducted in low- and middle-income countries among all articles (Spearman  $\rho$ , 0.51;  $P < .001$ ) and among articles reporting multicountry studies (Spearman  $\rho$ , 0.42;  $P = .005$ ) (**Figure 2**).

**Conclusions** The inclusion of editorial staff affiliated with low- and middle-income countries may be an approach to promoting the publication of research conducted in low- and middle-income countries.

<sup>1</sup>Precision Vaccines Program, Boston Children's Hospital, Boston, MA, USA, chris.rees@emory.edu; <sup>2</sup>Division of Pediatric Emergency Medicine, Emory University School of Medicine, Atlanta, GA, USA; <sup>3</sup>Children's Healthcare of Atlanta, Atlanta, GA, USA; <sup>4</sup>School of Public Health and Social Sciences, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; <sup>5</sup>Division of Infectious Diseases, Department of Paediatrics, The Hospital for Sick Children, Toronto, ON, Canada; <sup>6</sup>Department of Physiology

**Figure 2. Relationship Between Editorial Staff Composition From LMICs and the Proportion of Publications From LMICs**



LMICs indicates low- and middle-income countries.

and Nutrition, St. John's Medical College, Bengaluru, India; <sup>7</sup>Center for Nutrition, Division of Gastroenterology, Hepatology and Nutrition, Boston Children's Hospital, Boston, MA, USA; <sup>8</sup>Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** Christopher P. Duggan was supported in part by the National Institutes of Health (K24DK104676 and 2P30 DK040561).

**Role of the Funder/Sponsor** The funder had no role in the study design or in the collection, analysis, or interpretation of the data. The funder did not write the report and had no role in the decision to submit the abstract for presentation.

**Acknowledgments** Gandolina Melhem and Chris A. Rees are co-first authors. We thank Chloe Rotman, at Boston Children's Hospital, for her assistance in developing the PubMed query as well as acquiring the full-text articles reviewed in this study.

### Patterns of Gender and International Diversity of Editors and Editorial Boards Among Journals With Open Access Licenses and Open Science Policies

Micah Altman,<sup>1</sup> Philip Cohen<sup>2</sup>

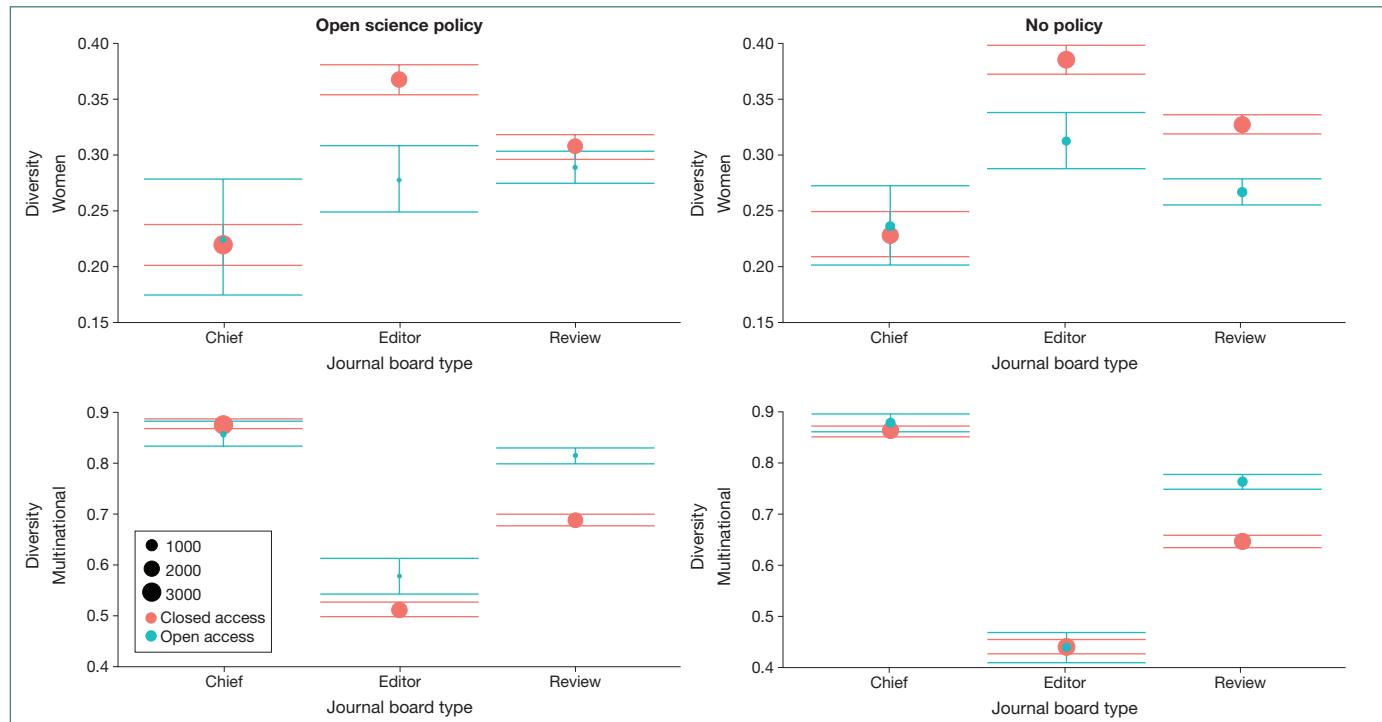
**Objective** This study measured diversity in scholarly journals' editors and editorial boards and evaluated associations between editor and board diversity and journal policies toward open access and open science. This study extends the literature on international diversity in editorial leadership positions<sup>1,2</sup> to a large multidisciplinary corpus of journals.

**Design** Multiple data sources were integrated to assemble a novel database describing the composition of editors and editorial boards, building on Pacher et al,<sup>3</sup> which used web mining to harvest editor-level information. These data were cleaned, coded, and supplemented with journal-level

classifications of journal discipline and journal policies derived from the 2018 Australian national research evaluation, Directory of Open Access Journals (DOAJ), and Center for Open Science TOPS initiative (TOPS). This dataset, collected in February 2021, comprised 17 publishers, 6090 journals, and 478,563 named editor roles. Diversity measures were computed for 14,228 journal editorial boards. Editor gender was imputed from editor names by comparison with Social Security Administration, Census, and social media corpora. Editor country was determined by applying geo-entity extraction and gazetteer lookup of organizational affiliation. Because chief editorship is often limited to a single post, diversity was measured across the pool of chief editors within that category. Boards were stratified by type (review, editor, chief) based on each journal's labeling of the board member's specific role. Journal policies were classified by access policy (ie, open or closed) based on whether the journal met DOAJ criteria (this corresponds approximately to gold and diamond open access categories) and by the presence of an open-science policy (applying the TOPS criteria). Means were calculated with bootstrapped 95% CIs.

**Results** The grand mean across board types was 28.7% for women representation and 69.7% for international diversity, as measured with the Herfindahl-Hirschman Index. There was substantial variation across disciplines: women composition of editorial boards ranged from 26% in engineering journals to 53% in education journals. **Figure 3** shows mean diversity by board type. Women composition of open access journal editorial boards was lower than that of closed-access journal editorial boards (-7%); in journals with open science policy vs those without (-4%), and in journals that were both open access and had an open science policy vs those without either (-6%). In contrast, open-access journal editorial boards were associated with higher international diversity (>11%) compared with closed-access journal boards. Furthermore, editorial boards were, on average,

**Figure 3. Diversity of Journal Boards**



Points indicate the diversity of the pooled category of chief editors and mean diversity of review and editorial boards. Bars indicate 95% CIs.

disproportionately male and USA-centric or UK-centric, although the specific level of diversity varied substantially by discipline.

**Conclusions** In this cross-sectional study of gender balance and international diversity in editorial boards, open-access journals were associated with lower women representation and more international diversity than their closed-access counterparts.

## References

1. Bhaumik S, Jagnoor J. Diversity in the editorial boards of global health journals. *BMJ Global Health*. 2019;4:e001909. doi:10.1136/bmjgh-2019-001909

2. Wu D, Lu X, Li J, Li J. Does the institutional diversity of editorial boards increase journal quality? the case economics field. *Scientometrics*. 2020;124:1579-1597. doi:10.1007/s11192-020-03505-6

3. Pacher A, Heck T, Schoch K. Open editors: a dataset of scholarly journals' editorial board positions. *SocArXiv*. Preprint published online March 10, 2021. Updated October 2, 2021. doi:10.31235/osf.io/jvzq7

<sup>1</sup>MIT Libraries, Massachusetts Institute of Technology, Cambridge, MA, USA, escience@mit.edu; <sup>2</sup>Department of Sociology, University of Maryland, College Park, MD, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** Supported in part by the Institute of Museum and Library Services, grant LG-250130-OLS-21.

## Factors Associated With Geographical Diversity of Reviewers Invited and Agreeing to Review for 21 Biomedical Journals

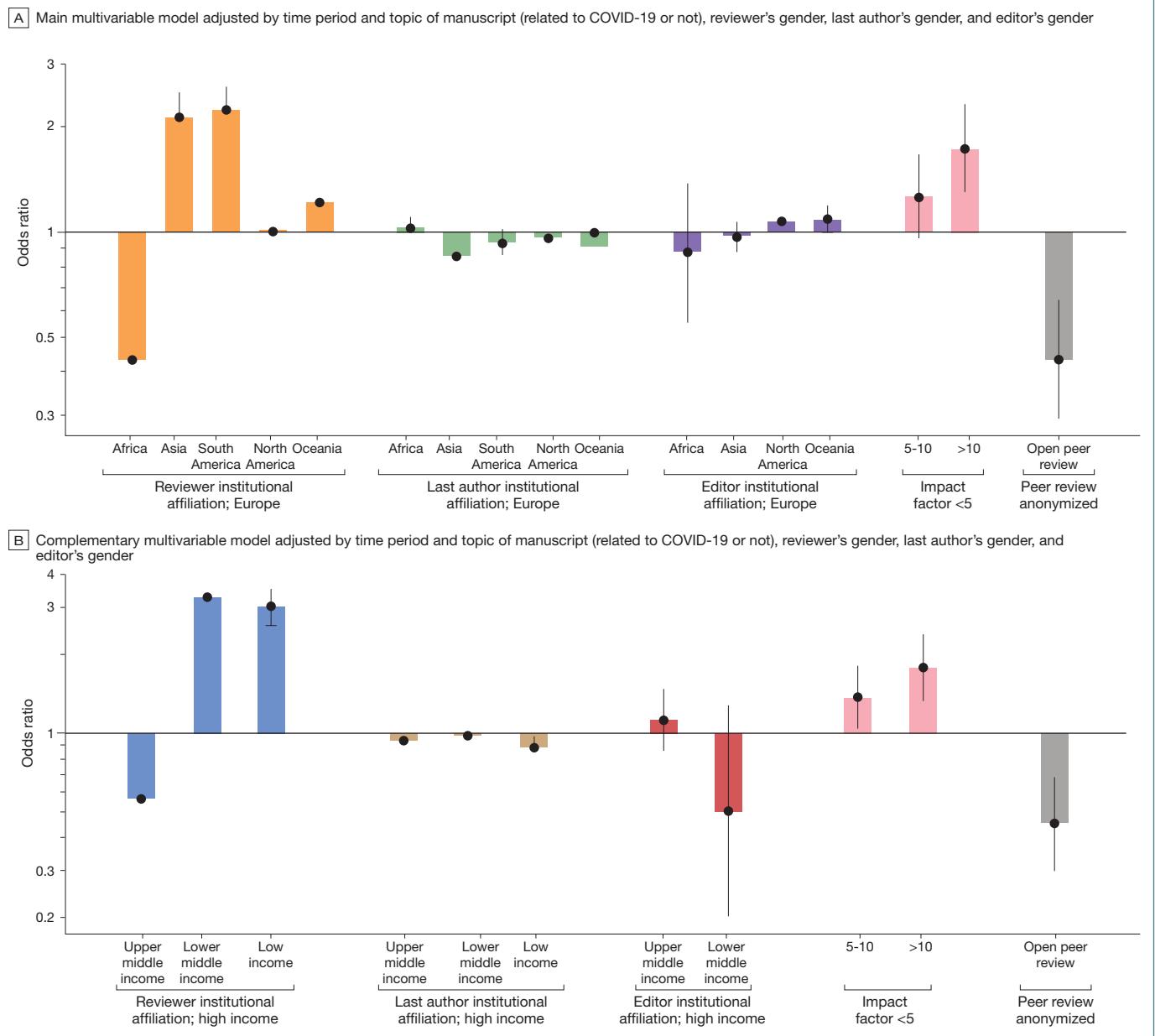
Khaoula Ben Messaoud,<sup>1,2</sup> Sara Schroter,<sup>3</sup> Mark Richards,<sup>4</sup> Angèle Gayet-Ageron<sup>1,2</sup>

**Objective** Geographical disparities have been observed in the acquisition of research grants and in the submission and publication of research articles. Peer reviewers are selected primarily based on their expertise through publication records. Response to peer invitation has been shown to be related to relevance of topic to own work, reviewer availability, journal attributes (impact factor and type of reviewer blinding) and compensation.<sup>1,2</sup> However, diversity in terms of geographical distribution in response to peer invitations remains almost unexplored.<sup>3</sup>

**Design** Retrospective cohort study of all research manuscripts submitted to 21 BMJ Publishing Group biomedical journals between January 1, 2018, and May 31, 2021, and subsequently sent for review. Data were collected on geographical affiliation, income level of the country of affiliation (according to World Bank Data 2020), journal impact factor, and peer-review process (open vs anonymized). The primary outcome was response (agreed vs not agreed) to review invitation. A multivariable mixed-effects logistic regression model with random factors on the intercept at journal and manuscript levels was performed.

**Results** A total of 257,025 reviewers were invited to review and 90,467 (35.2%) agreed. The distribution of geographical affiliations of the invited reviewers were as follows: 10.0% in Africa, 8.8% in Asia, 47.6% in Europe, 26.0% in North

**Figure 4. Likelihood of Agreeing to Review Manuscripts<sup>a</sup> Submitted to 21 BMJ Publishing Group Journals Between January 1, 2018, and May 31, 2021 (N=257,025) and Summary of Multivariable Models**



A, The main multivariable model adjusted for time period and topic of manuscript (related to COVID-19 or not), reviewer's gender, last author's gender, and editor's gender. B, The complementary multivariable model adjusted for time period and topic of manuscript (related to COVID-19 or not), reviewer's gender, last author's gender, and editor's gender. Models A and B were performed on complete observations ( $n = 194,908$  and  $n = 194,983$ , respectively). Mixed-effect logistic regression with manuscript and journal as random effects. The reference categories of the variables used in the model are specified horizontally below the bottom braces.

<sup>a</sup>Manuscripts with at least 1 review completed.

America, 6.3% in Oceania, and 1.2% in South America. Among invited reviewers, 217,682 (84.7%) were affiliated with a high-income country. **Figure 4** summarizes results from multivariable analysis. Agreement was higher among reviewers from Asia (2.13 [95% CI, 2.05-2.21]), Oceania (1.22 [95% CI, 1.17-1.27]), or South America (2.24 [95% CI, 2.06-2.45]) and lower among reviewers from Africa (0.43 [95% CI, 0.42-0.45] compared with Europe ( $P < .001$ ). Agreement was significantly lower when the last author was from Asia (0.85 [95% CI, 0.83-0.87]) or Oceania (0.91 [95% CI, 0.87-0.95]) compared with Europe ( $P < .001$ ). Reviewers agreed significantly more often when the associated editor

had a North American institutional affiliation compared with a European affiliation (1.07 [95% CI, 1.02-1.12]). Compared with high-income countries, agreement was higher among reviewers from lower middle-income countries (3.26 [95% CI, 3.06-3.48]) and low-income countries (2.99 [95% CI, 2.57-3.48]) ( $P < .001$ ). Agreement was also lower when the last author was from an upper middle-income country (0.94 [95% CI, 0.91-0.98]) or low-income country (0.88 [95% CI, 0.81-0.96]) compared with a high-income country ( $P < .001$ ). Agreement was associated with impact factor (higher for impact factors between 5 and 10, or >10, compared with <5: 1.73 [95% CI, 1.29-2.32] and lower when peer-review process

was open compared with anonymized: 0.43 [95% CI, 0.29–0.64]).

**Conclusions** The geographical affiliation of reviewers was an independent factor associated with agreement to review. To avoid bias and increase diversity, journal editors need to invite more reviewers from upper middle-income or low-income countries.

## References

1. Ellwanger JH, Chies JAB. We need to talk about peer-review: experienced reviewers are not endangered species, but they need motivation. *J Clin Epidemiol*. 2020;125:201-205. doi:10.1016/j.jclinepi.2020.02.001
2. Tite L, Schroter S. Why do peer reviewers decline to review? a survey. *J Epidemiol Community Health*. 2007;61(1):9-12. doi:10.1136/jech.2006.049817
3. Masukume G, Grech V. The Lancet peer reviewers: global pattern and distribution. *Lancet*. 2018;391(10140):2603-2604. doi:10.1016/S0140-6736(18)31136-X

<sup>1</sup>Department of Health and Community Medicine, University of Geneva, Geneva, Switzerland, Khaoula.BenMessaoud@unige.ch; <sup>2</sup>Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva, Switzerland; <sup>3</sup>The BMJ, London, UK; <sup>4</sup>Article Transfer Service, BMJ Publishing Group, London, UK

**Conflict of Interest Disclosures** All authors have completed the ICMJE uniform disclosure form at <https://www.icmje.org/disclosure-of-interest/> and declare support from the University Hospitals of Geneva, Geneva Medical School, and the Swiss National Science Foundation for the submitted work. Sara Schroter and Mark Richards are full-time employees of BMJ Publishing Group but were not involved in decision making on congress submissions. There were no other relationships or activities that could appear to have influenced the submitted work.

**Funding/Support** This study was supported by grant 192374 from the Swiss National Science Foundation.

**Role of the Funder/Sponsor** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the abstract.

## Comparison of Reporting Race and Ethnicity in Medical Journals Before and After Implementation of Reporting Guidance, 2019-2022

Annette Flanagin,<sup>1</sup> Miriam Y. Cintron,<sup>1</sup> Stacy L. Christiansen,<sup>1</sup> Tracy Frey,<sup>1</sup> Timothy Gray,<sup>1</sup> Iris Y. Lo,<sup>1</sup> Roger J. Lewis<sup>1,2,3,4</sup>

**Objective** Previous research found limited progress in reporting race and socioeconomic status in research published in medical journals from 2015 to 2019.<sup>1</sup> This study compared race and ethnicity reporting in 3 JAMA Network medical journals before and after implementation of Updated Guidance on Reporting Race and Ethnicity in Medical and Science Journals in August 2021.<sup>2</sup>

**Design** All major research articles published in *JAMA*, *JAMA Internal Medicine*, and *JAMA Pediatrics* in the first 3 months of 2019 and 2022 and the 3 months immediately before guidance implementation (May-July 2021) were included. Articles were reviewed independently by 2 reviewers for the following: study included human participants; race and ethnicity, age, sex and gender, and measures of socioeconomic status reported; where race and ethnicity was reported (abstract, methods, results, tables); number and order of racial and ethnic categories reported; if the category “other” was included and if that was defined; and if the article indicated how race and ethnicity were determined. Comparisons between years were calculated with  $\chi^2$  for 2-sided *P* values, odds ratios (ORs), and Wald test 95% CIs using R version 4.2.0.

**Results** Of 258 research articles published during the study periods, 249 (96.5%) included human participants and were included in this analysis. In 2019, 49 of 86 articles (57.0%) reported race and ethnicity of study participants compared with 42 of 77 (54.5%) in 2021 and 58 of 86 (67.4%) in 2022 (**Table 4**). Compared with articles reporting race and ethnicity, higher proportions of articles reported participants' age and sex or gender and a lower proportion reported socioeconomic status measures in all years. There were no significant differences in article location of reporting race and ethnicity or the proportion of articles that reported racial and ethnic categories. Of articles that included “other” as a collective racial and ethnic category, the proportion that defined specific categories included in “other” was 26.7% in 2019, 70.4% in 2021, and 84.8% in 2022, with a significant difference observed before guidance implementation in 2021 vs 2019 (OR, 6.53; 95% CI, 2.05-20.76; *P* < .001). A significant difference after reporting guidance implementation was observed for articles listing categories in alphabetical order (92.6% in 2022 vs 16.7% in 2021 [OR 73.75; 95% CI, 20.15-269.99; *P* < .001]) and articles indicating how race and ethnicity were determined (74.1% in 2022 vs 50.0% in 2021 [OR, 2.87; 95% CI, 1.23-6.66; *P* = .01]).

**Conclusions** In this analysis, higher proportions of articles reported how race and ethnicity were determined and listed categories in alphabetical order in 2022 following the implementation of reporting guidance. Some improvement was noted before 2021 and may have been associated with internal guidance shared in October 2020 and publication of an early draft of the guidance in February 2021.<sup>3</sup> Overall, race and ethnicity were still underreported compared with age, sex, and gender.

## References

1. Alegria M, Sud S, Steinberg BE, Gai N, Siddiqui A. Reporting of participant race, sex, and socioeconomic status in randomized clinical trials in general medical journals, 2015 vs 2019. *JAMA Netw Open*. 2021;4(5):e2111516. doi:10.1001/jamanetworkopen.2021.11516
2. Flanagin A, Frey T, Christiansen SL; AMA Manual of Style Committee. Updated guidance on the reporting of race and

**Table 4. Demographic Characteristics Reported in 3 Medical Journals From 2019 to 2022**

Reporting characteristics	2019		2021		2021 vs 2019		2022		2022 vs 2021	
	No.	%	No.	%	OR (95% CI)	P value	No.	%	OR (95% CI)	P value
Articles with human participants	86	95.6	77	96.3	NA	NA	86	97.7	NA	NA
Demographic characteristics reported										
Age	84	97.7	75	97.4	0.89 (0.12-6.50)	>.99	86	100	5.73 (0.27-121.20) <sup>a</sup>	.43
Sex and gender	79	91.9	73	94.8	1.62 (0.45-5.75)	.45	75	87.2	0.37 (0.11-1.23)	.09
Race and ethnicity	49	57.0	42	54.5	0.91 (0.49-1.68)	.76	58	67.4	1.73 (0.91-3.26)	.09
Socioeconomic status measures	32	37.2	34	44.2	1.33 (0.71-2.50)	.37	37	43.0	0.95 (0.51-1.78)	.88
Location of reporting race and ethnicity										
Abstract	16	18.6	18	23.4	1.33 (0.63-2.85)	.45	19	22.1	0.93 (0.45-1.94)	.85
Methods	37	43.0	41	53.2	1.51 (0.81-2.80)	.19	51	59.3	1.28 (0.69-2.38)	.44
Results	33	38.4	26	33.8	0.82 (0.43-1.56)	.54	40	46.5	1.71 (0.90-3.22)	.10
Table	45	52.3	40	51.9	0.98 (0.53-1.82)	.96	51	59.3	1.35 (0.72-2.51)	.35
Race and ethnicity categories										
Reported, median	5	NA	4.5	NA	NA	5	NA	NA	NA	
Articles reporting categories	48	55.8	42	54.5	0.95 (0.51-1.76)	.87	54	62.8	1.41 (0.75-2.63)	.28
Included “other” category	30	62.5	27	64.3	1.08 (0.46-2.55)	.86	33	61.1	0.87 (0.38-2.01)	.75
Defined categories included in “other”	8	26.7	19	70.4	6.53 (2.05-20.76)	<.001	28	84.8	2.36 (0.67-8.31)	.18
Presented categories in alphabetical order	4	8.3	7	16.7	2.20 (0.60-8.12)	.23	50	92.6	62.50 (17.00-229.84)	<.001
Indication how race and ethnicity determined	19	38.8	21	50.0	1.58 (0.69-3.64)	.28	43	74.1	2.87 (1.23-6.66)	.01

Abbreviations: NA, not applicable; OR, odds ratio.

<sup>a</sup>Calculated by adding 0.5 to each cell because of zero count.ethnicity in medical and science journals. *JAMA*.

2021;326(7):621-627. doi:10.1001/jama.2021.13304

3. Flanagin A, Frey T, Christiansen SL, Bauchner H. The reporting of race and ethnicity in medical and science journals: comments invited. *JAMA*. 2021;325(11):1049-1052. doi:10.1001/jama.2021.2104•JAMA Network, Chicago, IL, USA, annette.flanagin@jamamedical.org; <sup>a</sup>Berry Consultants LLC, Austin, TX, USA; <sup>b</sup>Harbor-UCLA Medical Center, Torrance, CA, USA; <sup>c</sup>David Geffen School of Medicine at UCLA, Los Angeles, CA, USA**Conflict of Interest Disclosures** Stacy L. Christiansen is chair and Annette Flanagin and Tracy Frey are members of the AMA Manual of Style committee. Annette Flanagin is executive director of the Peer Review Congress but was not involved in the review or decision for this abstract.

### Assessment of Neurology’s Implementation of Equity, Diversity, and Inclusion Editorial Review of Research Manuscripts

Roy H. Hamilton,<sup>1,2</sup> H. E. Hinson,<sup>1,3</sup> Rebecca Burch,<sup>1</sup> Joshua A. Budhu,<sup>1,4</sup> Nicole Rosendale,<sup>5</sup> Patricia K. Baskin,<sup>1</sup> Robert A. Gross,<sup>1,6,7</sup> José G. Merino<sup>1,8</sup>**Objective** In 2019, *Neurology* appointed 2 equity, diversity, and inclusion (EDI) editors to review manuscripts dealing with sex and gender, race and ethnicity, and other categories of marginalization.<sup>1,2</sup> Manuscripts requiring EDI review are identified by the EDI editor by title review or selected by the handling editor during the review process. The EDI review is

obtained only for manuscripts that are being considered for publication if suitably revised. Final editorial decisions are made considering the EDI reviews and the responses from the authors. The objectives of this retrospective analysis were to describe the outcomes of the process and highlight the issues identified by the EDI editors.

**Design** For this mixed methods study, the research articles that were reviewed by the EDI editors between May 1, 2019, and December 31, 2021, were identified. For this analysis, 75 of these manuscripts were selected at random. Two raters with expertise in EDI issues read the reviews and classified the issues raised by the EDI reviewers as related to terminology or conceptual issues in reference to sex and gender, race or ethnicity, or other marginalization categories. The analysis was limited to initial reviews. The data were summarized with descriptive statistics.**Results** In the study time frame, 7841 research articles were submitted to the journal. The EDI editors reviewed 101 manuscripts (1.3%). After evaluating the 75 reviews selected for analysis, 2 were excluded because the EDI editor reviewed them as a topic expert and not for EDI concerns. Among the 73 papers included in the analysis, the EDI reviewers identified at least 1 terminology or conceptual issue in 64, and there were at least 2 issues in 41 manuscripts. In relation to sex and gender, the EDI editors identified terminology and conceptual issues in 23 and 19 manuscripts, respectively; the matching numbers in relation to race or ethnicity were 23 and 56; and for other categories, 6 manuscripts had conceptual issues. The main themes identified during the EDI review process are presented in **Table 5**. As of May 31, 2022, 45 of

**Table 5. Main Issues Identified During the EDI Review Process**

Factual errors or misrepresentation of facts regarding marginalized groups
Insufficient methodological specificity regarding the nature of how EDI variables were conceptualized and determined
Insufficient acknowledgment of study limitations regarding EDI-related issues
Nonaffirming assertions and descriptions of marginalized groups (including obsolete terminology and biased language)
“Framing” issues (eg, implying that race is a biological rather than social construct)
Terminology concerns (eg, confusing sex and gender, race and ethnicity)

Abbreviation: EDI indicates equity, diversity, and inclusion.

the 73 manuscripts in the analysis had been accepted, 27 had been rejected, and 1 was undergoing major revision. The proportion of rejected manuscripts was lower for those in which an issue was identified by the EDI editor (22 of 64 vs 5 of 9; difference, 22%; 95% CI, −9.4% to 49.5%).

**Conclusions** The EDI editors identified several conceptual and nomenclature issues spanning various themes. The fact that the proportion of rejected manuscripts was lower when EDI concerns were identified suggests that the authors were able to address the concerns raised. The implementation of EDI review led to identification and substantive corrections related to categories of marginalization in manuscripts submitted to *Neurology*.

## References

1. Baskin PK, Pieper KM, Gross RA. Efforts by the journal to promote perspectives of equity, diversity, and inclusion and address disparities. *Neurology*. 2019;93(12):521-522. doi:10.1212/WNL.oooooooooooo0008021
2. Hamilton RH, Hinson HE. Introducing the associate editors for equity, diversity, and inclusion: aligning editorial leadership with core values in *Neurology®*. *Neurology*. 2019;93(15):651-652. doi:10.1212/WNL.oooooooooooo0008235

<sup>1</sup>*Neurology*, Minneapolis, MN, USA, jose.g.merino@gunet.georgetown.edu; <sup>2</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA; <sup>3</sup>Department of Neurology, Oregon Health and Science University, Portland, OR, USA; <sup>4</sup>Department of Neurology, Dana-Farber Cancer Institute, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; <sup>5</sup>Department of Neurology, University of California, San Francisco, CA, USA; <sup>6</sup>Department of Neurology, University of Rochester, Rochester, NY, USA; <sup>7</sup>Department of Pharmacology and Physiology, University of Rochester, Rochester, NY, USA; <sup>8</sup>Department of Neurology, Georgetown University Medical Center, Washington, DC, USA

**Conflict of Interest Disclosures** Roy H. Hamilton receives salary support from the American Academy of Neurology for his role as associate editor for Equity, Diversity, and Inclusion. H. E. Hinson receives salary support from the American Academy of Neurology for her role as associate editor for Equity, Diversity, and Inclusion. She consults for Biogen and conducts research supported by National Institute of Neurological Disorders and Stroke grant 1K23NS110828. Joshua A. Budhu receives salary support from the American Academy of Neurology for his role as section editor for Inclusion, Diversity, Equity, Anti-racism, and Social Justice. The Commonwealth Fund provides salary and research

support. Nicole Rosendale receives research support from the American Academy of Neurology. Patricia K. Baskin's salary is paid by the American Academy of Neurology. José G. Merino receives salary support from the American Academy of Neurology for his role as editor in chief of *Neurology*. No other disclosures were reported.

**Additional Information** Roy H. Hamilton and H. E. Hinson are co-first authors.

## Editorial and Peer Review Models

### Analysis of Submission Outcomes and Publication Timelines for Manuscripts Submitted to Cell Press Community Review Compared With Direct Journal Submissions

Sejal Vyas,<sup>1</sup> Matthew Pavlovich,<sup>1</sup> Jared Graves<sup>1</sup>

**Objective** Cell Press Community Review was launched in September 2020 with 20 participating journals as a pilot for a new peer review model. This model aimed to increase the efficiency of peer review by reducing cycles of editorial rejection through simultaneous consideration at multiple journals of interest and by providing delineated options for revisions for each target journal based on a single set of reviews. This study evaluated the performance of Community Review compared with the traditional peer review model, in which manuscripts may undergo multiple rounds of editorial rejection and transfer, to determine if Community Review results in improvements in the rates of peer review and acceptance and in publication timeline.

**Design** Total submissions, peer review offers, papers reviewed, and papers accepted from September 2020 through May 2022 were compared between Community Review and direct submissions to participating journals. Publication timeline for submissions to Community Review (Community Review submission date to journal acceptance date) were compared with those of papers submitted directly to the participating journals, delineated by the number of editorial reject/transfer cycles (first journal submission date to final journal acceptance date). Additionally, surveys were sent between April 2022 and May 2022 to authors on submission (38 responses of 258 sent, 15% participation rate) for feedback on the Community Review model.

**Results** Of 1674 total Community Review submissions, 1237 (74%) were offered peer review at 1 or more participating journals. Of the 1237 offers, 579 were taken up by the authors (47% uptake of peer review offer; 35% of total submissions reviewed). Of 344 Community Review submissions with a final postreview outcome, 191 were accepted (56% postreview accept rate). In the same period, 13,615 of 41,638 unique manuscripts (33%) directly submitted to participating journals were sent for review, either at the original journal or following 1 or more cycles of editorial reject/transfer. Of 9966 direct submissions with a final postreview outcome, 5689 were accepted (57% postreview accept rate). The results of the publication timeline analysis are presented in **Table 6**.

**Table 6. Results of Publication Timeline Analysis for Community Review Compared With Direct Journal Submission**

Peer review model	No. of editorial reject/transfer cycles	No. of papers	Mean publication timeline, d (SD)	Median publication timeline, d
Direct submission to participating journal	0	4158	174 (91)	158
	1	1356	204 (91)	185
	2	168	211 (89)	199
	3	6	249 (51)	234
	4	1	442 (NA)	NA
Community review	NA	156	201 (75)	192

Abbreviation: NA, not applicable.

Author surveys indicated “consideration across a selection of high-quality journals” and “interest in concept” as the top reasons authors chose to submit to Community Review.

**Conclusions** The Community Review model resulted in a greater chance of receiving an offer for peer review compared with direct journal submission and a larger percentage of total submissions being reviewed. The publication timeline for Community Review was comparable to that for direct journal submissions that underwent at least 1 round of editorial reject/transfer. There was interest from the research community in the Community Review submission model, with authors expressing interest in this model for efficiency in being considered for multiple journals simultaneously.

<sup>1</sup>Cell Press, Elsevier, Cambridge, MA, USA, svyas@cell.com

## Early Experiences of the Preprint Overlay Journal JMIRx

Gunther Eysenbach<sup>1</sup>

**Objective** In 2019, JMIR Publications announced the creation of a new series of journals, which are called superjournals, with *JMIRx-Med* the initial journal launched in that new journal series.<sup>1</sup>

Superjournals are a type of overlay journal that sit on top of preprint servers, offering rapid peer review and formal publication of revised preprints.<sup>1</sup>

**Design** An editorial prospecting platform (XDash) that invites preprint authors from medRxiv or bioRxiv was developed (**Figure 5**). All authors received an online survey that asked about motivations on why they posted the preprint, plans to submit to a journal, and whether they would be interested in a peer review of their preprint. Authors of new preprints were contacted as soon as their preprints were posted; no reminders were sent. Only preprints that were not already under peer review at a journal were eligible for consideration for *JMIRx*. After the peer review process, all peer review reports, author responses, and revised and accepted preprints are published in a *JMIRx* journal and deposited in PubMed Central/PubMed and institutional repositories of member institutions or, via the Manuscript

eXchange Marketplace, offered to other journals for publication.

**Results** Between December 19, 2019, and March 9, 2022, 11,143 responses were received. In the same period, 113,724 preprints were posted (response rate of 9.7%, although it is not known how many emails were actually extracted correctly and delivered). Forty-five percent of respondents (5011) submitted their manuscripts to a journal immediately after deposit as a preprint; for 1288 of respondents (11.6%), journals deposited a submitted manuscript on the preprint server on behalf of the authors. Together, these 2 groups represented 56.5% of respondents (n = 6299), whose preprint was already under peer review. The remaining 4844 respondents (43.5%) had not submitted their preprint to a journal yet. A total of 869 (7.8%) had no plans to submit to a journal, 3409 preprint authors (30.6%) planned to submit the preprint to a journal, and 1976 (17.7%) expressed interest in a peer review. A total of 7676 authors (68.9%) also indicated they absolutely need to publish the preprint in a journal with an impact factor, and 6560 authors (58.9%) indicated they wanted their publication indexed in PubMed.

**Conclusions** The *JMIRx* concept is welcomed by some authors looking for a rapid publication venue.

## References

- Eysenbach G. Celebrating 20 years of open access and innovation at JMIR Publications. *J Med Internet Res*. 2019;21(12):e17578. doi:10.2196/17578
- Stern BM, O’Shea EK. A proposal for the future of scientific publishing in the life sciences. *PLoS Biol*. 2019;17(2):e3000116. doi:10.1371/journal.pbio.3000116

<sup>1</sup>JMIR Publications, Toronto, ON, Canada, geysenba@gmail.com

**Conflict of Interest Disclosures** Gunther Eysenbach has an equity stake in and receives a salary from the publisher.

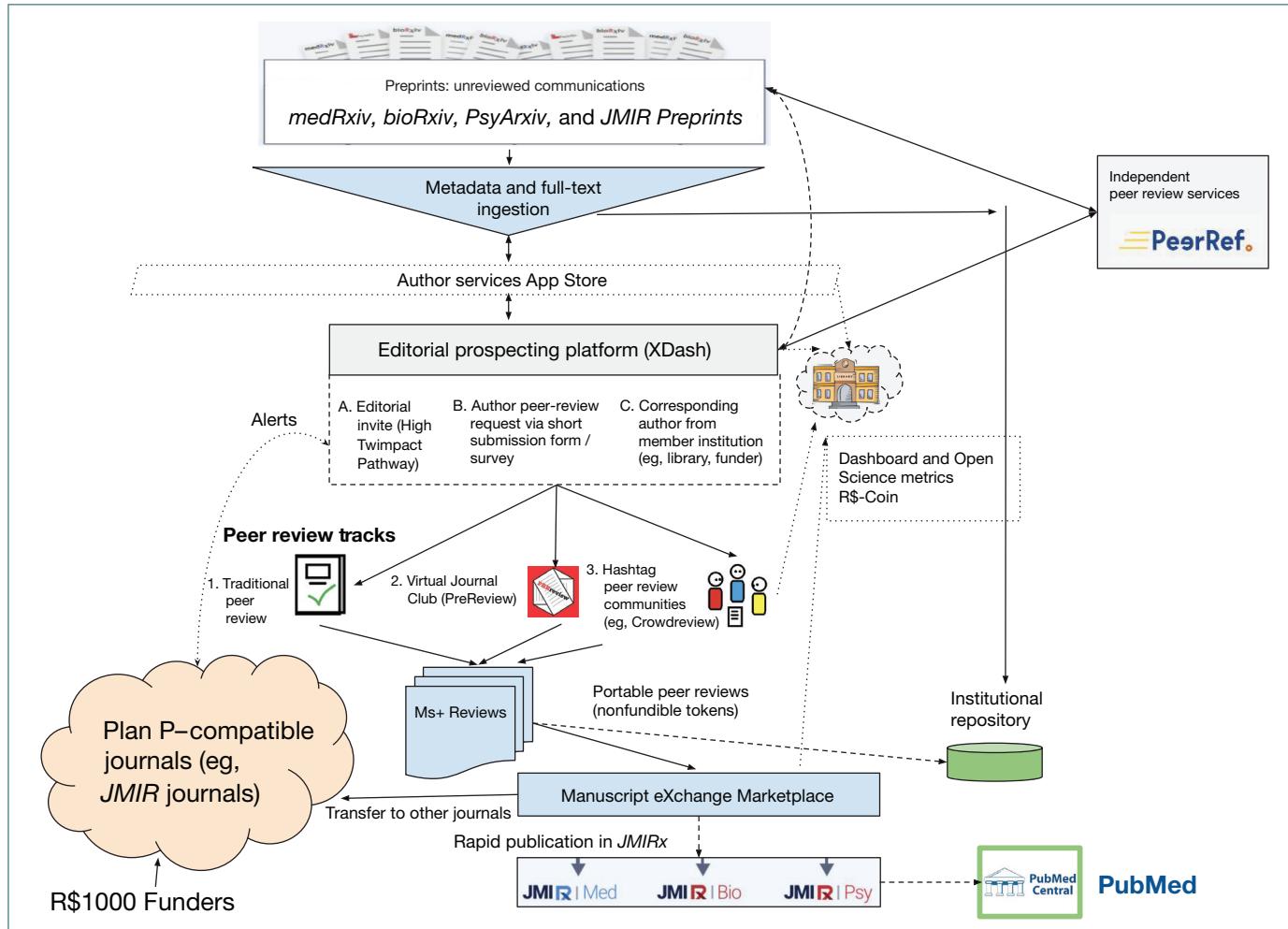
## Effect of Positive vs Negative Reviewer-Led Discussion on Herding and Acceptance Rates of Papers Submitted to a Machine Learning Conference: A Randomized Controlled Trial

Ivan Stelmakh,<sup>1</sup> Charvi Rastogi,<sup>1</sup> Nihar B. Shah,<sup>1</sup> Aarti Singh,<sup>1</sup> Hal Daumé III<sup>2,3</sup>

**Objective** Many publication venues and grant panels have a discussion stage in their peer review processes. Past research shows that human decisions in contexts involving social interactions (eg, financial markets) are susceptible to “herding,” ie, everyone doing what others are doing.<sup>1</sup> In this work, the possible presence of herding in peer review was investigated by testing whether reviewers are biased by the first argument presented in the discussion.

**Design** The experiment intervened in the peer review process of the 2020 International Conference on Machine Learning (ICML)—a large, top-tier machine learning conference—and executed a randomized controlled trial involving 1544 submissions and 2797 reviewers. In ICML,

**Figure 5. JMIRx Superjournal Workflow<sup>2</sup>**



discussion happens after reviewers submit their initial reviews. Based on these reviews, a set of borderline papers that had a disagreement between reviewers was chosen. These papers were uniformly at random split into 2 groups: positive and negative. An intervention was then implemented at the level of program chairs by sending the following requests. (1) Positive group: First, the most positive reviewer was asked to start the discussion. Later, the most negative reviewer was asked to contribute to the discussion. (2) Negative group: First, the most negative reviewer was asked to start the discussion. Later, the most positive reviewer was asked to contribute to the discussion. The goal of the intervention was to induce a difference in the order in which reviewers join the discussion across conditions. If herding was present, papers from the positive group were expected to have a higher acceptance rate than papers from the negative group. With this intervention, a permutation test at significance level  $\alpha = .05$  was used to compare final outcomes across groups (test statistic is the difference in acceptance rates). Importantly, it was empirically verified that the intervention did not result in confounding factors: all discussion parameters (other than the order) remained the same across groups.

**Results** Among 4625 ICML submissions, 1544 borderline papers (33%) were identified and allocated to the positive (755) and negative (789) groups. The intervention created a strong difference in the order in which reviewers joined the discussion (**Table 7**, rows 1 and 2), thereby confirming a strong detection power of the experiment. However, the difference in the order did not induce a difference in the outcomes (**Table 7**, row 3). Thus, there is no evidence of herding in peer review discussions of ICML 2020.

**Conclusions** Other applications involving discussions suffer from herding and, if present, herding could result in significant unfairness in acceptance of papers and awarding of grants. The finding of this study<sup>2</sup> was statistically negative but conveys a positive message: no evidence of herding is found in peer review, and hence no specific measures to counteract herding are needed.

## References

- Banerjee AV. A simple model of herd behavior. *Q J Econ*. 1992;107(3):797-817. doi:10.2307/2118364
- Stelmakh I, Rastogi C, Shah NB, Singh A, Daumé H III. A large scale randomized controlled trial on herding in peer-

**Table 7. Results of the Intervention**

	No. (%)		
	Positive group (n = 755 papers)	Negative group (n = 789 papers)	P value <sup>a</sup>
Most positive reviewer starts the discussion	400 (53)	71 (9)	.001
Most negative reviewer starts the discussion	113 (15)	466 (59)	.001
Acceptance rate	159 (21)	197 (25)	.12

<sup>a</sup>Values (2-sided) are for the difference between positive and negative groups.

review discussions. *arXiv*. Preprint posted online November 30, 2020. doi:10.48550/arXiv.2011.15083

<sup>1</sup>Carnegie Mellon University, Pittsburgh, PA, USA, stiv@cs.cmu.edu; <sup>2</sup>University of Maryland, College Park, MD, USA;

<sup>3</sup>Microsoft Research, New York, NY, USA

**Conflict of Interest Disclosures** Ivan Stelmakh reported having interned at Google; Charvi Rastogi reported having interned at IBM.

**Funding/Support** This work was supported by the US National Science Foundation (NSF): in part by NSF CAREER award 1942124, which supports research on the fundamentals of learning from people with applications to peer review, and in part by NSF CIF 1763734, which supports research on understanding multiple types of data from people.

**Role of the Funder/Sponsor** The NSF had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgements** We thank Edward Kennedy for useful comments on the initial design of this study. We are also grateful to the support team of the Microsoft Conference Management Toolkit (CMT) for their continuous support and help with multiple customization requests. Last, we appreciate the efforts of all reviewers and area chairs involved in the 2020 International Conference on Machine Learning review process.

**Additional Information** This study was approved by the Carnegie Mellon University Institutional Review Board.

## Pandemic Science

### Epidemiology of Scientific Output During the COVID-19 Pandemic

Anne Yang,<sup>1,2</sup> Jacob Kendall-Taylor,<sup>3</sup> Christopher C. Muth,<sup>3</sup> Jason N. Kennedy,<sup>2,4</sup> Stacy L. Christiansen,<sup>3</sup> Annette Flanagin,<sup>3</sup> Christopher W. Seymour<sup>2,3,4</sup>

**Objective** During the SARS-CoV-2 pandemic, clinicians, scientists, and public health experts raced to study the new disease, resulting in extraordinary scientific output. However, the relationship between pandemic scientific output and disease burden—particularly in characteristically underrepresented regions—is unknown.

**Design** This was a cohort study of research submissions to *JAMA* from May 1, 2020, to April 30, 2021, identified with a query “COVID-19” (n = 12,910) with a nonpandemic

comparison cohort of all submissions to *JAMA* from January 1 to December 31, 2019 (n = 9912). The following article types were included as “pandemic science”: clinical trial, meta-analysis, original investigation (ie, other full-length research submissions excluding the previous 2 categories), research letter, and brief report. Data on COVID-19 infections and deaths were sourced from World Health Organization (WHO) situation reports and analyzed by epidemiologically defined WHO-CHOICE subregions (Africa, Americas, Middle East, Europe, South-East Asia, and Western Pacific with China as a separate region).<sup>1,2</sup> WHO-CHOICE subregions are grouped A to E according to adult and child mortality, with subregion A having very low rates of adult and child mortality and subregion E having very high adult and child mortality.

**Results** In the pandemic cohort, the number of *JAMA* submissions reporting on COVID-19 were greater than the entirety of submissions in the nonpandemic cohort (12,910 vs 9912), particularly for original investigations (3449 vs 3152), research letters (2499 vs 823), and brief reports (1254 vs 495), respectively. However, this was not the case for submissions in the pandemic vs nonpandemic periods for clinical trials (317 vs 396) or meta-analyses (213 vs 311). In general, *JAMA* submissions tracked regionally with COVID-19 burden, with an increase in submissions several weeks after an increase in COVID-19 cases (**Figure 6**). When pandemic science was mapped according to WHO-CHOICE subregions, Americas A had the most submissions (4120), followed by Europe A (1944) and China (484). When measured as *JAMA* submissions per 1 million population, all subregions saw an increase during the pandemic period compared with the nonpandemic period, particularly in more developed subregions like Americas A (10.8 per 1 million in 2020 vs 8.1 per 1 million in 2019) and Europe A (4.3 per 1 million in 2020 vs 2.27 per 1 million in 2019).

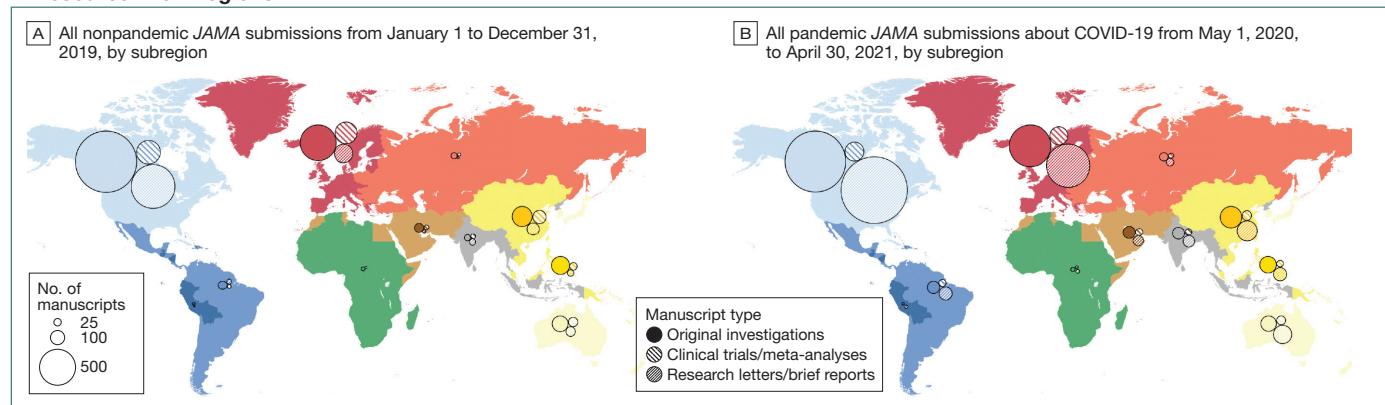
**Conclusions** Pandemic science submissions to *JAMA* from May 2020 to April 2021 exceeded all nonpandemic *JAMA* submissions in 2019. Although most submission types increased during the pandemic, there was a decrease in clinical trials and meta-analyses from all subregions. Resource-rich WHO-CHOICE subregions, such as Americas A and Europe A, produced the majority of COVID-19-related scientific output.

### References

- Evans DB, Edejer TT, Adam T, Lim SS. Methods to assess the costs and health effects of interventions for improving health in developing countries. *BMJ*. 2005;331(7525):1137-1140. doi:10.1136/bmj.331.7525.1137
- World Health Organization Novel Coronavirus (2019-nCoV) situation reports. Accessed June 1, 2021. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>

<sup>1</sup>Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA, yanga7@upmc.edu; <sup>2</sup>Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Pittsburgh, PA, USA; <sup>3</sup>JAMA Network, Chicago, IL, USA; <sup>4</sup>Department of Critical Care

**Figure 6. Geographic Distribution of Pandemic Science Submissions and WHO Choice Subregion With Greater Submissions (Circles) in Resource-Rich Regions**



During the pandemic period, there was an increase in original investigation, research letter, and brief report submissions and a decrease in clinical trials and meta-analyses.

Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

**Conflict of Interest Disclosures** Jacob Kendall-Taylor, Christopher C. Muth, Stacy L. Christiansen, Annette Flanagan, and Christopher W. Seymour are JAMA Network editors or editorial staff. Annette Flanagan is executive director of the Peer Review Congress but was not involved in the review or decision for this abstract.

### Comparison of the Characteristics of COVID-19 and Non–COVID-19 Retractions

Xiaoting Shi,<sup>1</sup> Alison Abrutis,<sup>2</sup> Rujvee P. Patel,<sup>3</sup> Mikas Grewal,<sup>4</sup> Ivan Oransky,<sup>2,5,6</sup> Joseph S. Ross,<sup>7,8,9</sup> Joshua D. Wallach<sup>1</sup>

**Objective** Concerns have been raised about the number of retracted COVID-19-related studies.<sup>1</sup> Although this may be due to the greater scrutiny of COVID-19-related literature, little is known about the potential differences between retracted COVID-19-related and non–COVID-19-related studies, including author characteristics and reasons for retractions.

**Design** In this cross-sectional analysis, all retractions of publications and withdrawals of preprints reporting the results of COVID-19-related and non–COVID-19-related medical studies indexed in the Retraction Watch Database between January 1, 2020, and May 5, 2022, were identified. Nonoriginal research (eg, retracted letters, viewpoints, and book chapters) and manuscripts classified by the Retraction Watch team as having been generated by companies that sell fake manuscripts (ie, paper mills) were excluded. For each retraction, the publication type (article or preprint) and design, number of authors, first and last authors and their affiliations, and date of retraction were recorded. The most prominent reasons for retraction were identified and grouped across similar categories. Scopus and then Google Scholar were searched to identify first and last author profiles, which were verified using affiliations, and we recorded an H-index and year of first publication for each author. The Fisher exact test and Mann-Whitney test were used to compare proportions of COVID-19-related vs non–COVID-19-related retractions.

**Results** Between January 1, 2020, and May 5, 2022, 140 COVID-19-related and 397 non–COVID-19-related studies were indexed in the Retraction Watch Database (**Table 8**); 72.1% of the COVID-19-related studies were peer-reviewed articles (27.9% were preprints), whereas 99.7% of the non–COVID-19-related studies were peer-reviewed articles. COVID-19-related studies were more likely to be retracted or withdrawn within 6 months of publication or posting than non–COVID-19-related studies (82.1% vs 58.2%;  $P < .001$ ). A greater proportion of modeling studies among the COVID-19-related than non–COVID-19-related studies was observed (13.6% vs 1.8%). COVID-19-related studies were more likely to be retracted without any explanation or to be removed for non–misconduct-related concerns than non–COVID-19-related studies (60.0% vs 35.5%). The first (29.1%) and last (30.7%) authors of COVID-19-related studies were more likely to have North American affiliations than the first (9.9%) and last (11.9%) authors of non–COVID-19-related studies. Nearly all first (97.6%) and last (98.0%) authors of all studies had academic or hospital affiliations. First and last authors of COVID-19-related studies had higher median (IQR) H-indexes than those of non–COVID-19-related manuscripts (7 [2-15] and 17 [4-27] vs 2 [1-6] and 5 [1-18]).

**Conclusions** Author and manuscript characteristics differed between retracted COVID-19-related and non–COVID-19-related studies. Although there have been hundreds of COVID-19-related retractions since the start of the pandemic, there have also been tens of thousands of preprints and published articles.

### Reference

1. Abrutis A, Marcus A, Oransky I. An “alarming” and “exceptionally high” rate of COVID-19 retractions? *Account Res.* 2021;28(1):58-59. doi:10.1080/08989621.2020.1793675

<sup>1</sup>Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA, joshua.wallach@yale.edu; <sup>2</sup>Retraction Watch, The Center for Scientific Integrity, New York, NY, USA; <sup>3</sup>Yale School of Public Health, New Haven, CT, USA; <sup>4</sup>Yale School of Medicine, New Haven, CT, USA; <sup>5</sup>Arthur Carter Journalism Institute, New York

**Table 8. Comparison Between COVID-19-Related and Non-COVID-19-Related Retractions**

Characteristic	Retractions, No. (%)			P value <sup>a</sup>	
	COVID-19 (n = 140)	Non-COVID-19 (n = 397)	Total (N = 537)		
<b>Manuscript characteristic</b>					
<b>Publication type</b>					
Peer-reviewed article	101 (72.1)	396 (99.7)	497 (92.6)	<.001	
Preprint	39 (27.9) <sup>b</sup>	1 (0.3) <sup>c</sup>	40 (7.4)		
Journal impact factor	n = 101	n = 396	n = 497		
Median (IQR)	3.070 (2.113-5.143)	3.432 (2.464-4.513)	3.355 (2.359-4.790)		
Journals without an impact factor	22 (21.8)	93 (23.5)	115 (23.1)		
<b>Study design</b>					
Trial	6 (4.3)	47 (11.8)	53 (9.9)		
Observational study	78 (55.7)	164 (41.3)	242 (45.1)		
Laboratory/preclinical	8 (5.7)	93 (23.4)	101 (18.8)		
Modeling	19 (13.6)	7 (1.8)	26 (4.8)		
Systematic review and/or meta-analysis	4 (2.9)	27 (6.8)	31 (5.8)		
Other	21 (15.0)	53 (13.4)	74 (13.8)		
Unclear	4 (2.9)	6 (1.5)	10 (1.9)		
<b>Reasons for retraction</b>					
Lack of adherence to policies/ethics	19 (13.6)	62 (15.6)	81 (15.1)	<.001	
Duplication of data, image, test, and/or article	12 (8.6)	55 (13.9)	67 (12.5)		
Errors	20 (14.3)	99 (24.9)	119 (22.2)		
Falsification and/or fabrication of data, images, and/or results	0	11 (2.8)	11 (2.0)		
Plagiarism of data, images, text, and/or article	5 (3.6)	29 (7.3)	34 (6.3)		
Non-misconduct-related other concerns, issues, or unspecified	84 (60.0)	141 (35.5)	225 (41.9)		
<b>Time from publication to retraction, mo</b>					
<6	115 (82.1)	231 (58.2)	346 (64.4)	<.001	
≥6	25 (17.9)	166 (41.8)	191 (35.6)		
<b>Author characteristic</b>					
Authors, median (IQR), No.	5 (3-9)	5 (3-7)	5 (3-8)	.50	
<b>First author</b>					
H-index					
Median (IQR)	7 (2-15)	2 (1-6)	3 (1-7)	<.001	
No information	7 (5.0)	31 (7.8)	38 (7.1)	.34	
Years since first publication					
Median (IQR)	9 (3-17)	4 (2-9)	5 (2-11)	<.001	
No information	7 (5.0)	31 (7.8)	38 (7.1)	.34	
<b>Last author</b>					
H-index					
Median (IQR)	17 (4-27)	5 (1-18)	7 (2-20)	<.001	
No information	5 (3.6)	15 (3.8)	20 (3.7)	>.99	
Years since first publication					
Median (IQR)	16 (9-25)	10 (3-19)	12 (3-20)	<.001	
No information	5 (3.6)	15 (3.8)	20 (3.7)	>.99	

<sup>a</sup>We used Fisher exact and Mann-Whitney tests to compare the characteristics of the COVID-19-related and non-COVID-19-related retractions.<sup>b</sup>medRxiv, n = 23; Social Science Research Network, n = 9; bioRxiv, n = 6; Research Square, n = 1.<sup>c</sup>Research Square, n = 1.

University, New York, NY, USA; <sup>6</sup>Spectrum, The Simons Foundation, New York, NY, USA; <sup>7</sup>Section of General Internal Medicine, Yale School of Medicine, New Haven, CT, USA; <sup>8</sup>Yale-New Haven Hospital Center for Outcomes Research and Evaluation, New Haven, CT, USA; <sup>9</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA

**Conflict of Interest Disclosures** Xiaoting Shi is supported by the China Scholarship Council in the past 36 months. Alison Abritis is an employee of The Center For Scientific Integrity, which developed and maintains the Retraction Watch Database. Ivan Oransky is the volunteer executive director of The Center For Scientific Integrity, a nonprofit organization that is funded through database licensing fees, a subcontract from the University of Illinois on a Howard Hughes Medical Institute grant, and donations from individuals. Joseph S. Ross is a former associate editor of *JAMA Internal Medicine*, a current research editor at *The BMJ*, and receives research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology (NEST), from the US Food and Drug Administration (FDA) for the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation program (U01FD005938), from the Agency for Healthcare Research and Quality (R01HS022882), from the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International; in addition, he is an expert witness at the request of relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. Joshua D. Wallach is supported by the FDA, Johnson & Johnson, and the National Institute on Alcohol Abuse and Alcoholism of the NIH under award 1K01AA028258.

### Comparison of Updates to Living Systematic Reviews Related to COVID-19 vs Other Subjects

Gustavo Magno Baldin Tiguman,<sup>1</sup> Marcus Tolentino Silva,<sup>2</sup> Taís Freire Galvão<sup>1</sup>

**Objective** Living systematic reviews (LSRs) are systematic reviews that are constantly updated to incorporate new evidence once it becomes available.<sup>1</sup> The aim of this study was to assess the frequency of updates to LSRs and the influence of COVID-19 on updates.

**Design** A cohort study of all LSRs indexed in PubMed, from inception to 2020, was performed, which were followed up until 2021. A PubMed search was conducted on September 2021 using the term *living systematic review* to retrieve articles published until December 31, 2020. Only the first version of the LSR was considered; protocols, conceptual papers, and other types of research were excluded. In January 2022, updates of included LSRs until December 31, 2021, were searched; the 2020 Journal Impact Factor (JIF) was collected from Journal Citation Reports; the open-access status of the articles (if publicly available in full text or not) was checked; and the LSR subject (COVID-19 or not) was assessed. The number of months from publication of the original LSR to its first update stratified by COVID-19 subject was calculated and a Kaplan-Meier survival analysis was performed. Regression analysis using Cox proportional

hazards models was conducted to calculate the hazard ratios (HRs) with 95% CIs adjusted by JIF and open-access status to minimize potential confounding.

**Results** Of 104 reports retrieved, 52 were original LSRs and were included in the study. Most LSRs were published in 2020 (33 of 52), and 19 were published between 2004 and 2019. Most articles were open access (50 of 52) and half had COVID-19 as the subject (28 of 52). As of December 31, 2021, a total of 60% (31 of 52) of the LSRs had been updated, taking a mean (SD) of 14 (3) months for the first update; 21 LSRs were not updated after a mean (SD) of 25 (4) months of publication. LSRs about COVID-19 were updated faster ( $n = 28$ ; mean [SD], 9 [1] months) than LSRs about other subjects ( $n = 24$ ; mean [SD], 30 [4] months; **Figure 7**). COVID-19 subject (HR, 3.64 [95% CI, 1.38-9.63];  $P = .009$ ) and higher JIF (HR, 1.10 [95% CI, 1.05-1.17];  $P < .001$ ) were associated with a higher probability of updating LSRs over time, while open-access status had no association with LSR updates (HR, 1.27 [95% CI, 0.16-10.25];  $P = .82$ ).

**Conclusions** This study identified a limited number of LSRs (52), and 60% of them were updated in about 1 year. Nearly 2 dozen were never updated despite having been published for 2 years, on average; future research may assess the reasons involved. The rapidly evolving COVID-19 pandemic and available research potentially favored more updated LSRs. LSRs published in high-impact journals were also more likely to be updated.

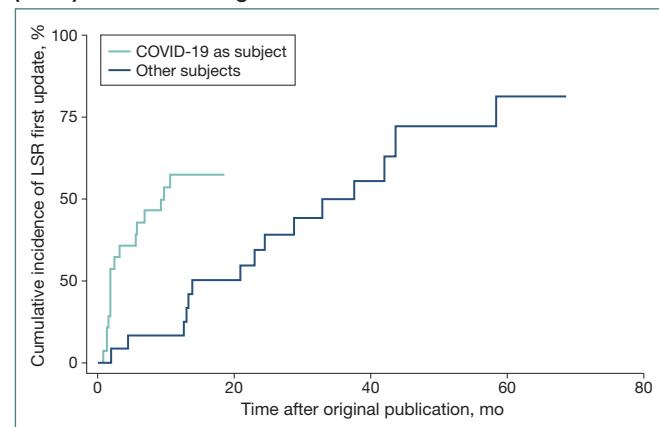
### Reference

- Elliott JH, Synnot A, Turner T, et al. Living systematic review: 1. introduction: the why, what, when, and how. *J Clin Epidemiol*. 2017;91:23-30. doi:10.1016/j.jclinepi.2017.08.010

<sup>1</sup>Faculty of Pharmaceutical Sciences, University of Campinas, Campinas, São Paulo, Brazil, taisgalvao@gmail.com; <sup>2</sup>Graduate Program of Pharmaceutical Sciences, University of Sorocaba, Sorocaba, Brazil

**Conflict of Interest Disclosures** None reported.

**Figure 7. Time to First Update of 52 Living Systematic Reviews (LSRs) Since Their Original Publication**



## Comparing Numerical Results Between Preprints and Peer-Reviewed Publications of COVID-19 Trials

Mauricia Davidson,<sup>1</sup> Anna Chaimani,<sup>1,2</sup> Isabelle Boutron<sup>1,2,3</sup>

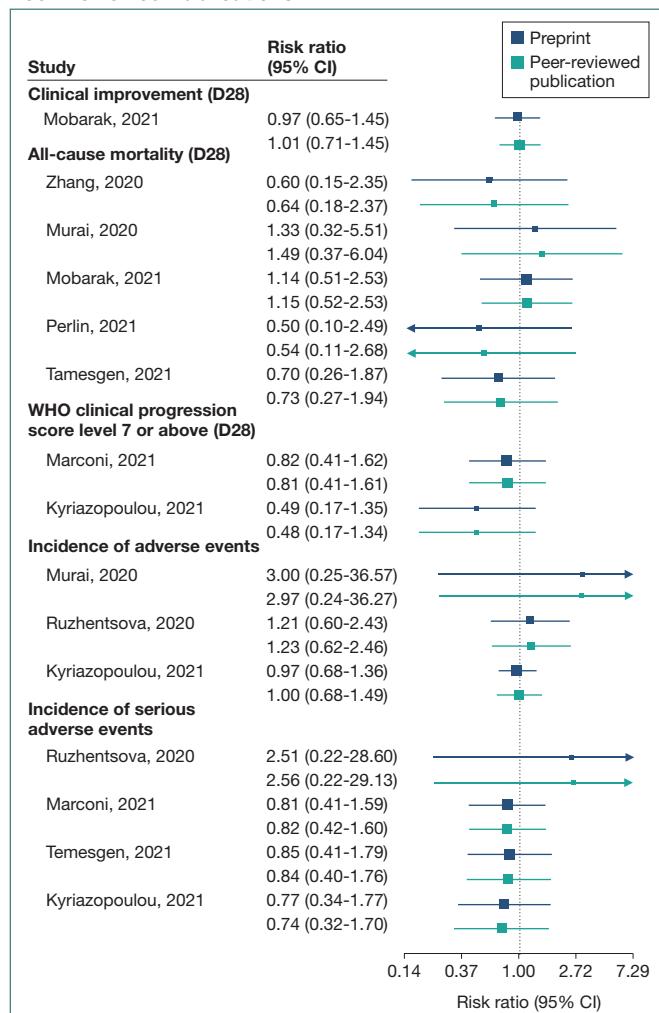
**Objective** The COVID-19 pandemic introduced a surge in the dissemination of preprints due to demand for faster and wider access to scientific knowledge. However, questions were raised concerning the reliability of their results.<sup>1,2</sup> The aim of this study was to compare numerical results extracted from preprints vs related peer-reviewed publications to inform inclusion in living systematic reviews.

**Design** This cross-sectional study used data from the COVID-NMA (covid-nma.com) initiative, a living systematic review of randomized clinical trials (RCTs) evaluating preventive interventions, treatments, and vaccines for COVID-19. Pharmacological treatment RCTs originally posted as preprints and subsequently published in peer-reviewed journals were included. Trials that moved from interim to final analysis between sources were excluded. Effect size estimates extracted from the first preprint were compared with effect size estimates from the most recent peer-reviewed publication. Predefined COVID-NMA “critical outcomes” at 28 days<sup>3</sup> were considered (ie, clinical improvement, World Health Organization Clinical Progression Score level 7 or above, all-cause mortality, incidence of any adverse events, incidence of serious adverse events). The last search date was February 3, 2022.

**Results** A total of 425 RCTs were identified. Trials only available as peer-reviewed publications ( $n = 217$  [51%]), preprints ( $n = 85$  [20%]), and unpublished ( $n = 16$  [4%]) were excluded, as well as trials reporting interim to final analysis between sources ( $n = 11$  [3%]), no review-specific outcomes ( $n = 4$  [1%]), and nonpharmacological treatments ( $n = 3$  [1%]). Eighty-nine RCTs (230 outcomes) first available as preprints and subsequently as peer-reviewed publications were included. The median delay between preprint post and subsequent publication in a peer-reviewed journal was 112 days (range, 5-505 days). Seventy-two (81%) preprint-publication RCTs (168 outcomes) showed no discrepancies in outcomes reported. Eight (9%) RCTs had numerical discrepancies in 15 of the 22 outcomes reported in both sources; no change in the direction of effect size estimate between sources was found (Figure 8). Of these, 1 RCT also had 2 outcomes added in the peer-reviewed publication. Furthermore, in trials with no numerical discrepancies in outcomes reported, 1 (1%) RCT had 2 outcomes missing in the peer-reviewed publication and 8 (9%) RCTs had at least 1 outcome added in the peer-reviewed publication compared with the preprint.

**Conclusions** Numerical results were generally similar between COVID-19 preprints and related peer-reviewed publications in the majority of RCTs. However, some outcomes were added and deleted. We could not assess whether preprint trials that were never published as peer-reviewed articles were problematic and whether peer review

**Figure 8. Discrepancy in Effect Size Estimates of Preprints vs Peer-Reviewed Publications**



WHO indicates World Health Organization.

prevented journal publication due to unsupported conclusions.

## References

- Flanigan A, Fontanarosa PB, Bauchner H. Preprints involving medical research—do the benefits outweigh the challenges? *JAMA*. 2020;324(18):1840-1843. doi:10.1001/jama.2020.20674
- Carneiro CFD, Queiroz VGS, Moulin TC, et al. Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *Res Integr Peer Rev*. 2020;5(1):16. doi:10.1186/s41073-020-00101-3
- Boutron I, Chaimani A, Devane D, et al. Interventions for the treatment of COVID-19: a living network meta-analysis. *Cochrane Database of Syst Rev*. Published online November 3, 2020. doi:10.1002/14651858.CD013770

<sup>1</sup>Université Paris Cité, INSERM, INRAE, CNAM, Centre of Research Epidemiology and Statistics, Paris, France, mauricia.davidson@etu.u-paris.fr; <sup>2</sup>Cochrane France, Paris, France; <sup>3</sup>Centre d’Epidémiologie Clinique, AP-HP, Hôpital Hôtel Dieu, Paris, France

**Conflict of Interest Disclosures** Mauricia Davidson is funded by a PhD grant from the Université Paris Cité. Isabelle Boutron is a

member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

**Funding/Support** This project is funded by the Université Paris Cité.

**Role of the Funder/Sponsor** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

### An Analysis of the History, Content, and Spin of Abstracts of COVID-19-Related Randomized Clinical Trials Posted as Preprints and Subsequently Published in Peer-Reviewed Journals or Unpublished

Hannah Spungen,<sup>1</sup> Jason Burton,<sup>2</sup> Stephen Schenkel,<sup>3</sup> David L. Schriger<sup>1</sup>

**Objective** Preprint servers have gained traction in many academic fields. Most preprints are unlikely to adversely affect public health. However, during a pandemic, when there is an urgent need for data, preprints may cause harm by disseminating incomplete, incorrect, or misleading information.<sup>1,2</sup> The aim of this study was to characterize and compare the characteristics, completeness, and spin of the abstracts of all randomized clinical trials (RCTs) related to COVID-19 posted to medRxiv from March 13, 2020, to December 31, 2021. An additional aim was to identify all corresponding published versions of these abstracts and perform a similar qualitative comparison to examine the impact of the peer review process.

**Design** An experienced librarian identified all COVID-19-related RCT preprints posted to medRxiv and all versions subsequently published in peer-reviewed journals as of June 1, 2022. An assistant created identically formatted Word documents of all abstracts. After training and the confirmation of adequate interrater reliability, 3 blinded reviewers scored individual abstracts presented in random order for completeness using items from CONSORT for Abstracts. They then evaluated blinded medRxiv/published abstract pairs for differences in content and spin, using criteria modified from Boutron et al.<sup>3</sup> Last, the abstracts of all unpublished preprints, along with an equal-sized sample of subsequently published preprints and their published counterparts, were assessed for extent of spin. Analysis was descriptive.

**Results** Two hundred ninety-one preprints were initially identified; 236 were confirmed as RCTs. Of these, 161 (68%) were found to have associated publications, which were published a median of 126 (IQR, 78-185; range, 0-654) days after medRxiv posting. The 75 unpublished preprints were posted a median of 344 (IQR, 235-429; range, 158-782) days prior to the final search. For most items, abstract completeness was higher in preprints that were subsequently published and was modestly higher still in published form (**Table 9**). The extent of spin was higher in unpublished preprints than in preprints that were subsequently published (**Table 9**). Last, of 161 published-preprint abstract pairs

studied, 25% had more spin in the preprint version, 8% had more spin in the published version, and 66% had no difference in spin. Conversely, 12% had more consensible preprints, 42% had more consensible published versions, and 45% had no difference.

**Conclusions** At this time, almost one-third of study medRxiv COVID-19-related RCTs have not been published, although roughly half were posted more than 1 year ago. This subset of unpublished preprints had lower CONSORT compliance and more spin than the medRxiv preprints that went on to be published. This study's comparison of published-preprint pairs is consistent with other literature that shows that peer review and the publication process improves—but does not eliminate—incomplete reporting and spin. Limitations of this study include the restriction of the search to a single preprint server and the focus on abstracts alone.

### References

1. Fraser N, Brierley L, Dey G, et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*. 2021;19(4):e3000959. doi:10.1371/journal.pbio.3000959
2. Kodvanj I, Homolak J, Virag D, et al. Publishing of COVID-19 preprints in peer-reviewed journals, preprinting trends, public discussion and quality issues. *Scientometrics*. 2022;127(3):1339-1352. doi:10.1007/s11192-021-04249-7
3. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303(20):2058-2064. doi:10.1001/jama.2010.651

<sup>1</sup>Department of Emergency Medicine, University of California, Los Angeles, CA, USA, hspungen@gmail.com; <sup>2</sup>UCLA Library, University of California, Los Angeles, CA, USA; <sup>3</sup>Department of Emergency Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

**Conflict of Interest Disclosures** None reported. David L. Schriger is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

**Funding/Support** David L. Schriger was supported in part by an unrestricted research grant from the Korein Foundation.

**Table 9. Characteristics, Completeness, and Spin in Blinded Preprint and Published Abstracts**

	No. (%) <sup>a</sup>				No. (%) <sup>a</sup>		
	Preprint abstracts, never published (n = 75)	Preprint abstracts, later published (n = 161)	Published abstracts (n = 161)		Preprint abstracts, never published (n = 75)	Preprint abstracts, later published (n = 161)	Published abstracts (n = 161)
<b>Abstract characteristics</b>							
No. of participants, median (IQR)	110 (46-353)	210 (89-809)	NA				
Journal impact factor, median (IQR)	NA	NA	14.8 (4.4-53.4)				
<b>Intervention type</b>							
Medication	50 (66.7)	106 (65.8)	NA				
Herbal or animal product	7 (9.3)	8 (5.0)	NA				
Vaccine	10 (13.3)	31 (19.3)	NA				
Behavioral	4 (5.3)	10 (6.2)	NA				
Nonpharmacologic	4 (5.3)	6 (3.7)	NA				
<b>Completeness<sup>b</sup></b>							
Randomization indicated in title	50 (66.7)	112 (69.6)	116 (72.1)				
General organization: abstract broken up into sections, eg, Results, Methods	66 (88.0)	140 (87.0)	127 (78.9)				
<b>Methods</b>							
Trial design described: eg, cluster randomized, parallel-group, superiority/noninferiority							
Fully	19 (25.3)	73 (45.3)	80 (50.0)				
Partially	52 (69.3)	80 (50.0)	74 (46.0)				
Clear objective/hypothesis given	39 (52.0)	106 (65.8)	116 (72.1)				
Participants	67 (89.3)	149 (92.5)	150 (93.2)				
Interventions	2 (96.0)	154 (95.7)	158 (98.1)				
Primary outcome(s) defined	38 (50.7)	113 (70.2)	122 (75.8)				
<b>Allocation to interventions described</b>							
Fully	2 (2.7)	9 (5.6)	21 (13.0)				
Partially	69 (92.0)	150 (93.2)	136 (84.5)				
Blinding discussed	52 (69.3)	117 (72.7)	121 (75.2)				
<b>Results</b>							
No. randomized	29 (38.7)	51 (31.7)	62 (38.6)				
<b>Recruitment (trial status)</b>							
Fully	15 (20.0)	47 (29.2)	59 (36.7)				
Partially	8 (10.7)	33 (20.5)	30 (18.6)				
No. analyzed in each group	21 (28.0)	61 (37.9)	73 (45.3)				
Results for primary outcome	23 (30.7)	86 (53.4)	93 (57.8)				
<b>Harms</b>							
Fully	8 (10.7)	27 (16.8)	35 (21.7)				
Partially	33 (44.0)	55 (34.2)	53 (32.9)				
Conclusion: general interpretation of results	73 (97.3)	154 (95.7)	155 (96.3)				
Trial registration listed	75 (100.0)	160 (99.4)	159 (98.8)				
Funding source listed	73 (97.3)	159 (98.8)	157 (97.5)				
<b>Presence of spin</b>							
(n = 75)	(n = 75) <sup>c</sup>	(n = 75) <sup>c</sup>	(n = 75) <sup>c</sup>				
Title	8 (10.7)	9 (12)	NA				
Results	16 (21.3)	2 (2.7)	NA				
Harms	7 (9.3)	2 (2.7)	NA				
Conclusions	49 (65.3)	30 (40.0)	NA				
<b>Extent of spin</b> , scored 0-10, median (IQR)	3 (1-6)	1 (0-3)	1 (0-2)				

Abbreviation: NA, not applicable.

<sup>a</sup>Unless otherwise specified.<sup>b</sup>Completeness items based on CONSORT for abstracts and spin items based on modified criteria from Boutron et al<sup>3</sup>; all preprints blinded prior to review.<sup>c</sup>Randomly selected pairs from all included preprints of subsequently published papers.

## Author and Peer Reviewer Guidance and Training

### Statistical Guidance to Authors at Top-Ranked Journals Across 22 Scientific Disciplines

Tom E. Hardwicke,<sup>1</sup> Maia Salholz-Hillel,<sup>2</sup> Mario Malički,<sup>3</sup> Denes Szűcs,<sup>4</sup> Theiss Bendixen,<sup>5</sup> John P. A. Ioannidis<sup>3,6,7</sup>

**Objective** Scientific journals may counter the misuse, misreporting, and misinterpretation of statistical methods by offering statistical guidance to authors.<sup>1,2,3</sup> This study assessed the nature and prevalence of statistical guidance in top-ranked journals across 22 scientific disciplines.

**Design** Statistical guidance from journal websites of 15 journals (top-ranked by impact factor) in each of 22 scientific disciplines (330 journals) was extracted and classified (in duplicate). Disagreements were resolved through discussion. Information was recorded on whether journals had dedicated statistical guidance sections and/or referred to guidance in external sources. For journals that provided their own statistical guidance, advice on each of 20 prespecified topics was recorded. For 6 topics that were considered in advance (based on author intuition) to be hotly debated in the statistical literature (*P* values, statistical significance, confidence intervals, effect sizes, sample size justification, and bayesian statistics), 1 investigator classified whether the

journal indicated opposition or endorsement and whether this was implicit or explicit.

**Results** Of 330 journals, 160 (48%) provided statistical guidance and 93 (28%) had a dedicated statistical guidance section in their author instructions (Figure 9, A). Statistical guidance was most common in health and life sciences journals. Notably, all 15 clinical medicine journals offered some statistical guidance. In 2 disciplines (computer science and mathematics), no journals offered any statistical guidance. Some journals shared the same publisher-level guidance, including 31 Nature Research journals (9%), 12 Cell Press journals (4%), and 2 Frontiers Media journals (0.6%). A total of 137 journals (42%) referred authors to statistical guidance in 80 individual external sources, 49 of which were reporting guidelines (the remainder were primarily journal articles). Among 20 prespecified statistical topics (Figure 9, B), only 2 were mentioned in more than a quarter of the journals: confidence intervals (90 [27%]) and *P* values (88 [27%]). Guidance on these topics was inconsistent across journals. For 6 hotly debated topics, only 3 journals explicitly opposed the use of statistical significance; more commonly, journals implicitly endorsed the use of *P* values (77 [23%]), statistical significance (35 [11%]), and bayesian statistics (39 [12%]) and explicitly endorsed reporting of effect sizes (62 [19%]), confidence intervals (85 [26%]), and sample size justifications (67 [20%]).

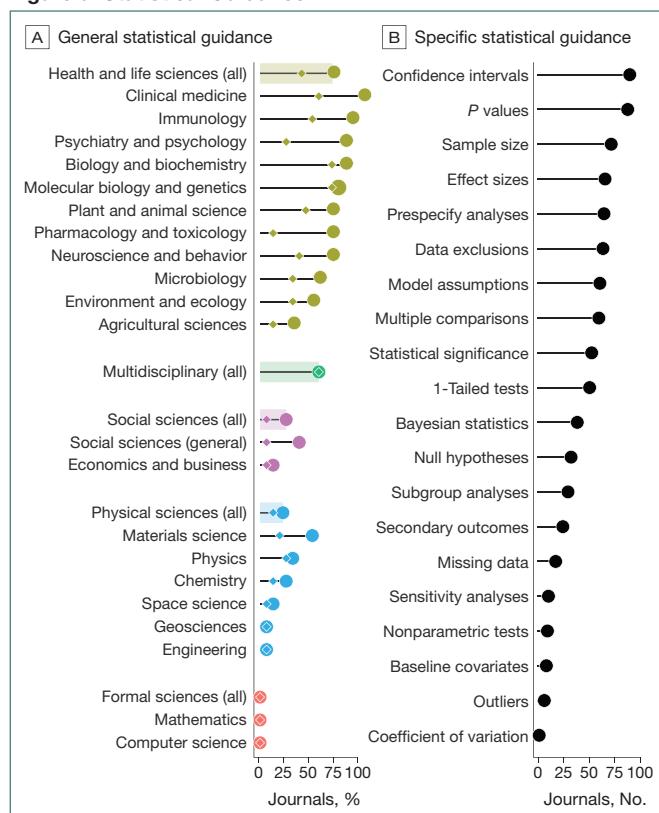
**Conclusions** The results of this study suggest that there are large gaps and inconsistent coverage in the statistical guidance provided by top-ranked journals across scientific disciplines. Future studies should investigate whether journal statistical guidance to authors is associated with improved selection, use, reporting, or interpretation of statistical analyses.

### References

1. Strasak AM, Zaman Q, Marinell G, Pfeiffer K, Ulmer H. The use of statistics in medical research: a comparison of the *New England Journal of Medicine* and *Nature Medicine*. *Am Statistician*. 2007;61:47-55. doi:10.1198/000313007X170242
2. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J (Clin Res Ed)*. 1983;286:1489-1493.
3. Schriger DL, Arora S, Altman DG. The content of medical journal instructions for authors. *Ann Emerg Med*. 2006;48:743-749. doi:10.1016/j.annemergmed.2006.03.028

<sup>1</sup>Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands, hardwickete@gmail.com; <sup>2</sup>QUEST Center for Responsible Research, Berlin Institute of Health at Charité—Universitätsmedizin Berlin, Berlin, Germany; <sup>3</sup>Meta-Research Innovation Center at Stanford, Stanford University, Palo Alto, CA, USA; <sup>4</sup>Department of Psychology, University of Cambridge, Cambridge, UK; <sup>5</sup>Department of the Study of Religion, Aarhus University, Aarhus, Denmark; <sup>6</sup>Meta-Research Innovation Center Berlin, QUEST Center for Transforming Biomedical Research, Berlin Institute of Health at Charité—Universitätsmedizin Berlin, Berlin, Germany; <sup>7</sup>Departments of Medicine, Epidemiology and

Figure 9. Statistical Guidance



**Conflict of Interest Disclosures** Tom E. Hardwicke receives funding from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant 841188. Maia Salholz-Hillel is employed as a researcher under research grants from the German Bundesministerium für Bildung und Forschung. Mario Malicki is a co-editor in chief of *Research Integrity and Peer Review*. John P. A. Ioannidis is a codirector of the Peer Review Congress but was not involved in the review or decision of this abstract. The Meta-Research Innovation Center at Stanford is supported by a grant from the Laura and John Arnold Foundation, and the Meta-Research Innovation Center Berlin is supported by a grant from the Einstein Foundation and Stiftung Charité. No other disclosures were reported.

## Reminding Peer Reviewers of the Most Important Reporting Guideline Items to Improve Completeness in Published Articles: Primary Results of 2 Randomized Controlled Trials

Benjamin Speich,<sup>1,2</sup> Erika Mann,<sup>3</sup> Christof M. Schönenberger,<sup>2</sup> Katie Mellor,<sup>1</sup> Alexandra N. Griessbach,<sup>2</sup> Paula Dhiman,<sup>1,4</sup> Pooja Gandhi,<sup>5,6</sup> Szimonetta Lohner,<sup>7,8</sup> Arnav Agarwal,<sup>9,10</sup> Ayodele Odutayo,<sup>1,11</sup> Iratxe Puebla,<sup>12</sup> Alejandra Clark,<sup>12</sup> An-Wen Chan,<sup>13</sup> Michael M. Schlussel,<sup>14</sup> Philippe Ravaud,<sup>14,15</sup> David Moher,<sup>16,17</sup> Matthias Briel,<sup>2,9</sup> Isabelle Boutron,<sup>14,15</sup> Sara Schroter,<sup>18</sup> Sally Hopewell<sup>1</sup>

**Objective** Reporting guidelines have been available since 1994. Numerous studies have shown that adherence to reporting guidelines is suboptimal,<sup>1,2</sup> raising the question of whether a specific targeted intervention for peer reviewers might improve reporting. The aim of this study was to evaluate whether asking peer reviewers, via email, to check if specific reporting guideline items were adequately reported in the submitted manuscripts they were reviewing would improve adherence to reporting guidelines in published articles.

**Design** Two parallel-group superiority randomized controlled trials (RCT-1 and RCT-2) using submitted manuscripts as the unit of randomization. RCT-1 focused on RCT protocols and how well they were reported considering the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) guidelines, and RCT-2 focused on RCT results publications and the reporting of CONSORT (Consolidated Standards of Reporting Trials) items. Manuscripts in both RCTs were randomized (1:1) to intervention or control; the control group received usual journal practice. RCT-1 included manuscripts containing RCT protocols submitted from June 2020 to May 2021 to *BMJ Open* that were sent for peer review (<https://osf.io/z2hm9>). The RCT-2 trial<sup>3</sup> included manuscripts describing RCT primary results, submitted from July 2019 to July 2021 to 1 of 7 journals (5 BMJ Publishing Group; 2 Public Library of Science [PLOS]). In the intervention group (both trials), peer reviewers received an email from the journal reminding them to check if items were adequately reported in the manuscript. In RCT-1, these were the 10 most important and poorly reported SPIRIT items and for RCT-2, the 10 most important

and poorly reported CONSORT items. In both RCTs, peer reviewers and authors were not informed of the purpose of the study and outcome assessors were blinded. The primary outcome was the difference in the mean proportion of adequately reported 10 SPIRIT and 10 CONSORT items between intervention and control in the final published article.

**Results** In RCT-1, 245 manuscripts were randomized. Of those, 178 were published (90 intervention; 88 control). A mean proportion of 46.1% (95% CI, 41.8%–50.4%) of the 10 SPIRIT items were adequately reported in the intervention group and 45.6% (95% CI, 41.7%–49.4%) in the control group (mean difference, 0.5%; 95% CI, −5.2% to 6.3%) (**Figure 10**). In RCT-2, of the 511 randomized manuscripts, 243 were published (121 intervention; 122 control). A total of 67.4% (95% CI, 63.8%–71.1%) of the 10 CONSORT items were adequately reported in the intervention group and 65.9% (95% CI, 61.9%–69.9%) in the control group (mean difference, 1.5%; 95% CI, −3.8% to 6.9%) (**Figure 10**).

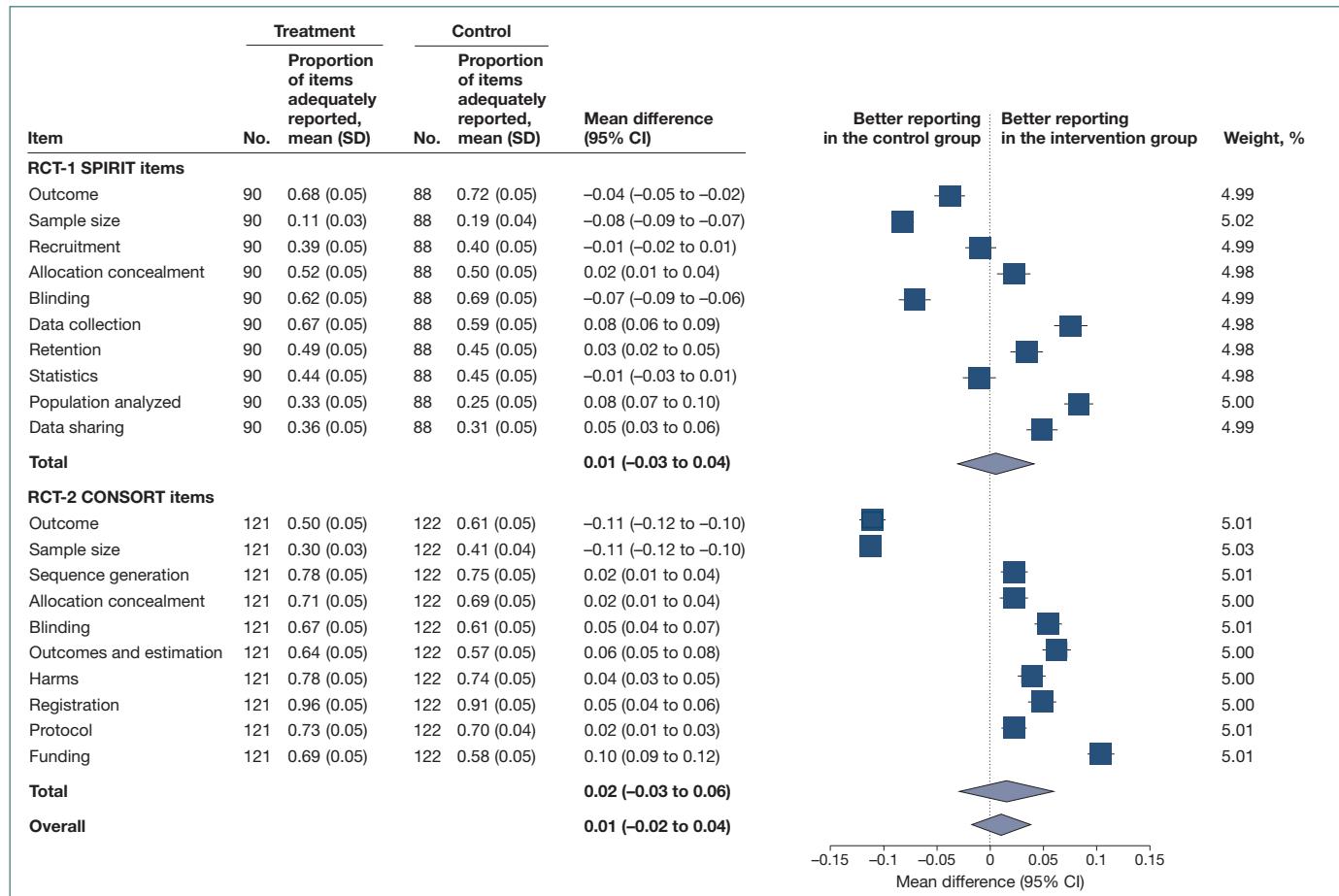
**Conclusions** Journals asking peer reviewers, via email, to check if the most important and poorly reported items are adequately reported in submitted manuscripts did not improve the reporting completeness of the final published article.

## References

1. Samaan Z, Mbuagbaw L, Kosa D, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *J Multidiscip Healthc*. 2013;6:169–188.
2. Jin Y, Sanger N, Shams I, et al. Does the medical literature remain inadequately described despite having reporting guidelines for 21 years? a systematic review of reviews: an update. *J Multidiscip Healthc*. 2018;11:495–510.
3. Speich B, Schroter S, Briel M, et al. Impact of a short version of the CONSORT checklist for peer reviewers to improve the reporting of randomised controlled trials published in biomedical journals: study protocol for a randomised controlled trial. *BMJ Open*. 2020;10(3):eo35114.

<sup>1</sup>Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK; <sup>2</sup>Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland, benjamin.speich@usb.ch; <sup>3</sup>PLOS, Public Library of Science, San Francisco, CA, USA; <sup>4</sup>The EQUATOR Network, Oxford, UK; <sup>5</sup>Rehabilitation Sciences Institute, University of Toronto, Toronto, ON, Canada; <sup>6</sup>Swallowing Rehabilitation Research Laboratory, Toronto Rehabilitation Institute—University Health Network, Toronto, ON, Canada; <sup>7</sup>Cochrane Hungary, Clinical Centre of the University of Pécs, Medical School, University of Pécs, Pécs, Hungary; <sup>8</sup>Department of Public Health Medicine, Medical School, University of Pécs, Pécs, Hungary; <sup>9</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada; <sup>10</sup>Division of General Internal Medicine, Department of Medicine, McMaster University, Hamilton, ON, Canada; <sup>11</sup>Applied Health Research Centre, Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, ON, Canada; <sup>12</sup>PLOS ONE, Public Library of Science, San Francisco, CA, USA, and Cambridge, UK; <sup>13</sup>Department of Medicine, Women's College Research Institute,

**Figure 10. Difference in the Mean Proportion of Adequately Reported SPIRIT and CONSORT Items**



The values in this plot differ from the values in the abstract because the test in the meta-analysis is based on a different concept than the *t* test used for primary analysis.

Women's College Hospital, University of Toronto, Toronto, ON, Canada; <sup>14</sup>Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique Hôpitaux de Paris (APHP) Paris, France; <sup>15</sup>Université de Paris, CRESS, INSERM, INRA, Paris, France; <sup>16</sup>Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada; <sup>17</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>18</sup>*The BMJ*, London, UK

**Conflict of Interest Disclosures** Erika Mann and Alejandra Clark are employed by the Public Library of Science. Paula Dhiman, Michael M. Schlussel, Philippe Ravaud, and David Moher are members of the EQUATOR (Enhancing the Quality and Transparency of Research) network. During the design and initial implementation, Iratxe Puebla was an employee by the Public Library of Science. An-Wen Chan and David Moher are authors of the SPIRIT 2013 Statement (Standard Protocol Items: Recommendations for Interventional Trials). David Moher, Sally Hopewell, and Isabelle Boutron are members of the Consolidated Standards for Reporting Trials (CONSORT) group and authors of the CONSORT 2010 statement. Sara Schroter is employed by *The BMJ*. David Moher is an associate director and An-Wen Chan and Isabelle Boutron are advisory board members of the International Congress on Peer Review and Scientific Publication but were not involved in the review or decision for this abstract. No other disclosures were reported.

**Funding/Support** This study was supported in part by the Swiss National Science Foundation.

**Additional Information** The study was registered on the Open Science Framework (<https://osf.io/c4hn8>).

### Assessment of a Structured and Mentored Peer Review Curriculum on Quality of Peer Review

Ariel M. Lyons-Warren,<sup>1,2</sup> Whitley W. Aamodt,<sup>3</sup> Roy Stroud,<sup>4</sup> Kathleen M. Pieper,<sup>5</sup> José G. Merino<sup>5,6</sup>

**Objective** The Resident & Fellow Section (RFS) of *Neurology* expanded and formalized a biannual virtual, mentored peer review training program in 2020. Similar to prior studies,<sup>1</sup> qualitative data demonstrated that mentored peer review improved understanding of and confidence with independent review.<sup>2</sup> The objective of the current study was to quantitatively evaluate review quality before and after a mentored peer review program for neurology residents and fellows.

**Design** In this pre-post intervention study, faculty mentors chosen from a national pool of experienced reviewers with an interest in mentoring were paired with trainee peer reviewers. Mentees were selected from residents and fellows who responded to a call for program participants via American Academy of Neurology social media channels or who applied to the RFS editorial board. Mentees first completed unassisted reviews of a standardized manuscript. Participants

**Table 10. Modified RQI Scores Before and After Program Participation**

Modified RQI		Score range	Preprogram score, mean (SD) (n = 20)	Postprogram score, mean (SD) (n = 18)	<i>P</i> value
Question No.	Topic				
1	Summary of study	0-1	0.77 (0.42)	1.00 (0)	.04
2	Separate comments		0.30 (0.47)	0.91 (0.32)	<.001
3	Organization		0.62 (0.45)	0.61 (0.42)	.50
4	Recommendation		0.48 (0.50)	0.91 (0.24)	.002
5	Importance	1-5	1.93 (0.86)	2.70 (1.25)	.07
6	Originality		1.90 (0.71)	2.48 (1.13)	.02
7	Method		2.75 (1.16)	3.18 (1.15)	.02
8	Writing		2.55 (0.96)	2.94 (1.17)	.18
9	Constructive		3.03 (0.99)	3.21 (1.03)	.07
10	Use examples		2.73 (1.16)	2.70 (1.09)	.20
11	Results		2.33 (0.96)	2.82 (0.96)	.009
12	References		1.28 (0.52)	1.58 (0.58)	.63
13	Tone		3.57 (0.62)	3.73 (0.45)	.32
14	Overall quality		2.67 (1.07)	2.94 (1.13)	.03
NA	Total	10-55	26.38 (6.65)	31.70 (6.69)	<.001

Abbreviations: NA, not applicable; RQI, Review Quality Index.

then received structured resources on how to review a manuscript and construct written peer review. Mentor-mentee dyads completed 2 reviews over 6 months. Mentees also completed unassisted postprogram reviews of a standardized manuscript. The association between program participation and peer review quality was quantitatively assessed by 3 independent evaluators scoring the preprogram and postprogram unassisted reviews using a modified version of the Review Quality Instrument (RQI)<sup>3</sup> comprising 4 items scored present or absent (1 or 0) and 10 items scored on a 5-point Likert scale for a maximum score of 55. Higher scores indicated better review quality. Evaluator scores were averaged for each participant, and a paired *t* test was used to compare preintervention and postintervention scores. *P* < .05 was considered significant.

**Results** A total of 20 mentor-mentee pairs, including 8 incoming members of the RFS editorial board and 12 neurology trainees, were enrolled over 2 sessions in 2021. Two mentees failed to complete the postprogram unassisted review and were excluded. Total modified RQI scores ranged from 15.67 to 39.00 before the program and 19.67 to 43.67 after the program. The mean total score for all participants increased following completion of the program (preprogram, 26.38; postprogram, 31.70; *P* < .001). Postprogram reviews were more likely to include separate comments for editors and authors (mean preprogram score, 0.30; mean postprogram score, 0.91; *P* < .001), more likely to identify the strengths and weaknesses of the research methods (mean preprogram score, 2.75; mean postprogram score, 3.18; *P* = .02), and more likely to comment on the interpretation of results (mean preprogram score, 2.33; mean postprogram score, 2.82; *P* = .009). Preprogram and postprogram scores

for all 14 measures of the modified RQI are shown in **Table 10**.

**Conclusions** Peer review quality improved for residents and trainees following completion of a structured, mentored peer review curriculum. Formal mentoring to teach the proper approach to peer review is one tool to expand the bench of available quality peer reviewers.

## References

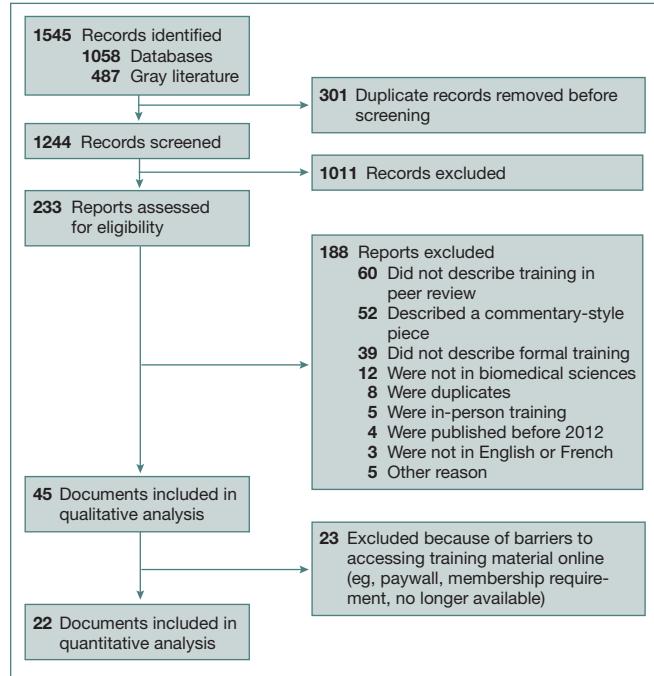
- Wong VSS, Strowd RE 3rd, Aragón-García R, et al. Mentored peer review of standardized manuscripts as a teaching tool for residents: a pilot randomized controlled multi-center study. *Res Integr Peer Rev.* 2017;2:6. doi:10.1186/s41073-017-0032-0
- Lyons-Warren AM, Aamodt WW, Pieper KM, Strowd R. A sustainable, comprehensive, mentored program for teaching peer review to neurology residents. Presented at: American Academy of Neurology Annual Meeting, April 17-22, 2021.
- van Rooyen S, Black N, Godlee F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. *J Clin Epidemiol.* 1999;52(7):625-629. doi:10.1016/S0895-4356(99)00047-5

<sup>1</sup>Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA, lyonswar@bcm.edu; <sup>2</sup>Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA; <sup>3</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA; <sup>4</sup>Wake Forest School of Medicine, Winston-Salem, NC, USA; <sup>5</sup>Neurology, Pittsford, NY, USA; <sup>6</sup>Department of Neurology, Georgetown University Medical Center, Washington, DC, USA

**Conflict of Interest Disclosures** Ariel M. Lyons-Warren reported serving as the chair of the Resident and Fellow Section Mentored Review Program. Whitley W. Aamodt reported receiving an editorial stipend from *Neurology* and serving as the deputy section editor for the Resident and Fellow section of *Neurology*. Roy Strowd reported serving as a consultant for Monteris Medical Inc and Novocure; receiving an editorial stipend as section editor for the Resident and Fellow section of *Neurology*; having received stipends as an educational lecturer for Lecturio and Kaplan; and having received research and grant support from the American Academy of Neurology (AAN), the American Board of Psychiatry and Neurology, American Society for Clinical Oncology, Southeastern Brain Tumor Foundation, and Jazz Pharmaceuticals. Kathleen M. Pieper reported receiving salary support from AAN and serving as the senior managing editor of *Neurology*. José G. Merino reported receiving salary support from the AAN for his editorial role and serving as editor in chief of *Neurology*.

**Acknowledgements** We thank the residents and fellows who participated in the program and completed the preprogram and postprogram assessments as well as the senior reviewers who served as mentors. In addition, we thank Seth Retzlaff in the *Neurology* editorial office for the handling and tracking of manuscripts. We greatly appreciate their time and commitment to improving the next generation of peer reviewers.

**Figure 11. PRISMA Flow Diagram**



## Online Training in Scholarly Peer Review: A Systematic Review

Jess V. Willis,<sup>1,2</sup> Janina Ramos,<sup>1,3</sup> Ryan Chow,<sup>1,2</sup> Mohsen Alayche,<sup>1,2</sup> Jeremy Y. Ng,<sup>1</sup> Kelly D. Cobey,<sup>1,4</sup> David Moher<sup>1,5</sup>

**Objective** To perform a systematic review of available online training for scholarly peer review of biomedical journal articles.

**Design** A search strategy was developed and reviewed using the PRESS (Peer Review of Electronic Search Strategies) checklist by a medical librarian. A database search of MEDLINE, PsycINFO, Embase, ERIC, and Web of Science was conducted. Additional supplementary searches were done of preprint servers, Google, YouTube, university library websites, publisher websites, and peer review-related events and groups. All English or French training documents for scholarly peer review of biomedical manuscripts freely accessible online between January 1, 2012, and the date of the search (September 13, 2021) were included. January 1, 2012, was used as the earliest cutoff because this was the year Publons was launched. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram with full exclusion criteria is shown in **Figure 11**. Screening was done in duplicate by 2 independent reviewers in 2 stages, with conflicts resolved by a third party. Data extraction and risk of bias were done by 1 reviewer and then verified by a second. As no current risk of bias tool could be found for evaluating training material, one was created, which was pilot tested for feasibility.

**Results** Of 1244 records screened, 45 online training documents were identified for peer review. Barriers such as paywalls and membership requirements limited access to just more than half of these documents (23 of 45 [51%]); thus, they were excluded from data analysis because they were not

freely available online. The included documents were mostly websites (13 of 22 [59%]) and videos (6 of 22 [27%]) offered exclusively in English (19 of 22 [86%]). Many of the documents did not report a creation year (10 of 22 [45%]), author information (10 of 22 [45%]), or funding sources (19 of 22 [86%]). Countries that developed the greatest amount of training were the US (8 of 22 [36%]), United Kingdom (4 of 22 [18%]), and Germany (3 of 22 [14%]). The main training formats were online modules (13 of 22 [59%]) and webinars (5 of 22 [23%]) and took less than 1 hour to complete (15 of 22 [68%]). Topics that were frequently included were an overview of the peer review process (18 of 22 [82%]), synthesis of a peer review report (20 of 22 [91%]), and critical appraisal of data (18 of 22 [82%]). Conversely, critical appraisal of clinical trials (4 of 22 [18%]), statistics (4 of 18 [3%]), and reporting guidelines (9 of 22 [41%]) were less commonly included.

**Conclusions** This systematic review identified a comprehensive list of available online training material for scholarly peer review of biomedical journals and an analysis of their characteristics.

<sup>1</sup>Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada, jwill135@uottawa.ca; <sup>2</sup>Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>3</sup>Department of Biology, Faculty of Science, University of Ottawa, Ottawa, ON, Canada;

<sup>4</sup>University of Ottawa Heart Institute, Ottawa, ON, Canada; <sup>5</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

**Conflict of Interest Disclosures** David Moher is an associate director of the Peer Review Congress but was not involved in the review or decision of this abstract.

## Peer Review

### Development of a Global Dataset for Peer Review in Astronomy

Vicente Amado Olivo,<sup>1</sup> Wolfgang Kerzendorf<sup>2</sup>

**Objective** The great astronomical observatories accept thousands of proposals per year from astronomers hoping to receive telescope time. Specifically, the Space Telescope Science Institute receives approximately 1000 proposals per year for the Hubble Space Telescope, with this number projected to double as the James Webb Space Telescope has safely launched.<sup>2</sup> In astronomy, a Time Allocation Committee (TAC) reviews all proposals submitted for the use of a telescope and identifies the proper expert to review the proposal. The goal of the study was to develop a database of all active astronomers and their publications that assists in the identification of experts for the peer review of observing proposals, expanding on work done by Kerzendorf et al<sup>1</sup> and Strolger et al.<sup>2</sup>

**Design** The database creation and modeling study has expanded the reviewer pool to all around the world, instead of simply relying on the TAC's personal networks. The Semantic Scholar Open Research Corpus (S2ORC) data set allowed for

the creation of a preliminary database consisting of authors, their full-text publications, and associated metadata. The identification of experts for peer review was systematically done by leveraging an astronomer's body of work (ie, scientific publications). An author's publications and the observing proposal were numerically represented using machine learning models to identify which astronomer's expertise is similar for review of the proposal. Various methods were compared to disambiguate author names using name-based techniques. However, authors with full names having more than 3 words were excluded owing to formatting issues (currently investigating methods to address the issue). A preliminary prototype using machine learning and natural language processing models was tested using 918 proposals from the European Southern Observatory (significant metrics to evaluate expertise are being researched).

**Results** The S2ORC data set, which consists of 12 million full-text publications, was filtered to only astronomy publications using publication arXiv identifiers. The database contains 212,839 publications and a total of 1,801,916 nonunique authors from 1991 to 2020. Three author name disambiguation algorithms were compared: first initial, all initials, and hybrid method.<sup>3</sup> The 3 methods were validated using an initial subset of 1538 ORCID identifiers matched to astronomers. A contamination rate is the percentage of validated astronomers whose identity became compromised due to merging or splitting of names. The contamination rates of the 3 methods were 1.77%, 15.52%, and 2.02%, respectively.

**Conclusions** The developed database has expanded the possible reviewer pool from several hundreds known to the TAC to all active astronomers worldwide. A larger pool of reviewers allows for more accurate expertise matching.

## References

1. Kerzendorf WE, Patat F, Bordelon D, van de Ven G, Pritchard TA. Distributed peer review enhanced with natural language processing and machine learning. *Nat Astron.* 2020;4(7):711-717. doi:10.1038/s41550-020-1038-y
2. Strolger LG, Porter S, Lagerstrom J, Weissman S, Reid IN, Garcia M. The Proposal Auto-Categorizer and Manager for time allocation review at Space Telescope Science Institute. *AJ.* 2017;153(4):181. doi:10.3847/1538-3881/aa6112
3. Milojević S. Accuracy of simple, initials-based methods for author name disambiguation. *J Informetrics.* 2013;7(4):767-773. doi:10.1016/j.joi.2013.06.006

<sup>1</sup>Michigan State University, East Lansing, MI, USA, amadovic@msu.edu

**Conflict of Interest Disclosures** None reported.

**Funding/Support** Funding was received from the Space Telescope Science Institute–Hubble Space Telescope Science Policies Group.

**Role of Funder/Sponsor** The funder collaborated with members from the Space Telescope Science Institute to oversee the development of the database.

---

## Comparison of Review Scores of Computer Science Conference Submissions With Cited and Uncited Reviewers

Charvi Rastogi,<sup>1</sup> Ivan Stelmakh,<sup>1</sup> Ryan Liu,<sup>1</sup> Shuchi Chawla,<sup>3</sup> Federico Echenique,<sup>3</sup> Nihar B. Shah<sup>1</sup>

**Objective** Many anecdotes suggest that including citations to the works of potential reviewers is a good (albeit unethical) way to increase the acceptance chances of a paper. However, previous attempts<sup>1,2</sup> to quantify this effect (citation bias) had low sample sizes and unaccounted confounding factors, such as paper quality (stronger papers had longer bibliographies) or reviewer expertise (cited reviewers had higher expertise). In this work, the question of whether positive comments from reviewers are associated with their work being cited in the papers that they review was investigated.

**Design** The study used data from 2 top-tier computer science conferences: the 2021 Association for Computing Machinery Conference on Economics and Computation (EC) and 2020 International Conference on Machine Learning (ICML). Both conferences received full-length papers that underwent rigorous review (similar to top journals in other areas). The study analyzed anonymized observational data, and consent collection was not required. The dependent variable of the analysis was the overall score given by a reviewer to a paper (between 1 and 5 in EC and 1 and 6 in ICML; higher meant better). To investigate the association between the citation of a reviewer and their score, parametric (linear regression for EC and ICML) and nonparametric (permutation test with covariate matching for ICML) tests at significance level  $\alpha = .05$  were combined, circumventing various confounding factors, such as paper quality, genuinely missing citations, reviewer expertise, reviewer seniority, and reviewers' preferences in which papers to review. The approach comprised matching cited and uncited reviewers within each paper and then carefully analyzing the differences in their scores. In this way, the aforementioned paper quality confounder was alleviated as matched cited and uncited reviewers reviewed the same paper. Additionally, various attributes of reviewers (eg, their expertise in the paper's research area) were used to account for confounders associated with the reviewer identity (eg, reviewer expertise). Finally, the genuinely missing citation confounder was accounted for by excluding papers in which an uncited reviewer genuinely decreased their evaluation of a paper because it failed to cite their own relevant past work.

**Results** Overall, 3 analyses were conducted, with sample sizes ranging from 60 to 1031 papers and from 120 to 2757 reviewers' evaluations. These analyses detected citation bias in both venues and indicated that citation of a reviewer was associated with an increase in their score (approximately 0.23 point on a 5-point scale). For reference, a 1-point increase of a score by a single reviewer would improve the position of a paper by 11% on average.

**Conclusions** To improve peer review, it is important to understand the biases present and their magnitude. This

work<sup>3</sup> studied citation bias and raised an important open problem of mitigating the bias. The reader should be aware of the observational nature of this study when interpreting the results.

## References

1. Beverly R, Allman M. Findings and implications from data mining the IMC review process. *ACM SIGCOMM Computer Communication Review*. 2012;43(1):22-29. doi:10.1145/2427036.2427040
2. Sugimoto CR, Cronin B. Citations gamesmanship: testing for evidence of ego bias in peer review. *Scientometrics*. 2013;95(3):851-862. doi:10.1007/s11192-012-0845-z
3. Stelmakh I, Rastogi C, Liu R, Echenique F, Chawla S, Shah NB. Cite-seeing and reviewing: a study on citation bias in peer review. *arXiv*. Preprint posted online March 31, 2022. doi:10.48550/arXiv.2203.17239

<sup>1</sup>Carnegie Mellon University, Pittsburgh, PA, USA, crastogi@cs.cmu.edu; <sup>2</sup>The University of Texas at Austin, Austin, TX, USA; <sup>3</sup>California Institute of Technology, Pasadena, CA, USA

**Conflict of Interest Disclosures** Ivan Stehlmakh reported interning for Google. Charvi Rastogi reported interning for IBM. No other disclosures were reported.

**Funding/Support** Ivan Stehlmakh, Charvi Rastogi, Ryan Liu, and Nihar B. Shah were supported by a National Science Foundation (NSF) CAREER award (1942124), which supports research on the fundamentals of learning from people with applications to peer review. Federico Echenique was supported by NSF grants SES 1558757 and CNS 1518941. Charvi Rastogi was partially supported by a J. P. Morgan artificial intelligence research fellowship.

**Role of the Funder/Sponsor** The funders did not play a role in any of the following: design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgement** We appreciate the efforts of all reviewers involved in the review process of the 2021 Association for Computing Machinery Conference on Economics and Computation and 2020 International Conference on Machine Learning. We thank Valerie Ventura, PhD, Carnegie Mellon University, for useful comments on the design of the analysis procedure.

**Additional Information** Nihar B. Shah is a co-corresponding author.

## Association Between Author Prominence and Peer Reviewers' Willingness to Review and Their Evaluations of Manuscripts Submitted to a Finance Journal

Jürgen Huber,<sup>1</sup> Sabiou Inoua,<sup>2</sup> Rudolf Kerschbamer,<sup>3</sup> Christian König-Kersting,<sup>1</sup> Stefan Palan,<sup>4</sup> Vernon L. Smith<sup>2</sup>

**Objective** Merton<sup>1</sup> argued that “eminent scientists get disproportionately great credit for their contribution to science while relatively unknown scientists tend to get disproportionately little credit for comparable contributions.” In this context, this study asked (1) is there a status bias in reviewers’ propensity to accept review invitations? and (2) is there a status bias in their evaluation of the paper?

**Design** A manuscript written by Vernon Smith (Nobel laureate, high prominence) and Sabiou Inoua (young researcher, low prominence) was submitted to the *Journal of Behavioral and Experimental Finance* for peer review. The paper was assigned to 3299 reviewers randomized into 5 conditions: (1) no author name in the invitation email or on the manuscript’s title page (treatment: AA; 576 reviewers); (2) high-prominence author name on manuscript only (AH; 696); (3) low-prominence author name on manuscript only (AL; 739); (4) high-prominence author name on both email and manuscript (HH; 507); and (5) low-prominence author name on both email and manuscript (LL; 781). To avoid confounding, only 1 name was shown in the email and on the manuscript, and the author was always designated as the corresponding author. Reviewers gave consent to being part of the study prior to accessing the paper. Those who submitted a report were debriefed after the study. Reviewers’ decisions to accept the invitation in response to anonymized (AA, AH, AL) vs nonanonymized (LL, HH) emails were compared using Fisher exact tests. The distribution of publication recommendations (eg, reject, major revision, minor revision, or accept) was compared for manuscripts that showed the author’s name (AL, AH) vs those that did not (AA) using Mann-Whitney tests.

**Results** A total of 2611 researchers (79.1%) responded to the invitation, 821 of whom agreed to review (31.4%). The invitation showing Vernon Smith was accepted statistically significantly more often than those showing no author name or Sabiou Inoua (acceptances: HH, 158 of 410 [38.5%] vs LL, 174 of 610 [28.5%];  $P = .001$ ; HH, 158 of 410 [38.5%] vs anonymized, 489 of 1591 [30.7%];  $P = .003$ ). Of the 821 reviewers who accepted the invitation, 534 (65.0%) submitted reports (AA, 110; AL, 101; AH, 102; LL, 114; and HH, 107). The manuscript showing the prominent author received 53.3% less reject recommendations and more than 10 times as many accept recommendations as the anonymized version (test on the distribution of recommendations: AH vs AA,  $P < .001$ ) (**Table 11, A**). The manuscript showing the name of the less prominent author got 35.5% more reject recommendations and 63.7% less minor revision recommendations than the anonymized version (test on the distribution of recommendations: AL vs AA,  $P = .005$ ) (**Table 11, B**). Author prominence affected the willingness to review and reviewers’ recommendations.

**Conclusions** Although double-anonymized peer review is not a panacea,<sup>2</sup> this study’s results still support its use in the field of finance.

## References

1. Merton, RK. The Matthew effect in science. *Science*. 1968;159(3810):56-63. doi:10.1126/science.159.3810.56
2. Snodgrass, R. Single- versus double-blind reviewing: an analysis of the literature. *ACM SIGMOD Record*. 2006;35(3):8-21. doi:10.1145/1168092.1168094

<sup>1</sup>Department of Banking and Finance, University of Innsbruck, Innsbruck, Austria, christian.koenig@uibk.ac.at;

**Table 11. Results Overview**

Panel A: willingness to review <sup>a</sup>					
	Anonymized <sup>b</sup> (AA, AL, AH)	Low <sup>c</sup> (LL)	High (HH)	Total	
Invitations sent, No.	2011	781	507	3299	
Responses received, No.	1591	610	410	2611	
Invitations accepted, No.	489	174	158	821	
Acceptance rate, %	30.7	28.5	38.5	31.4	
Panel B: reviewers' recommendations <sup>d</sup>					
Condition, %	Reject	Major revision	Minor revision	Accept	Total No.
AL <sup>e</sup>	65.3	24.8	7.9	2.0	101
AA	48.2	28.2	21.8	1.8	110
AH <sup>f</sup>	22.5	18.6	38.2	20.6	102

Abbreviations: AA, no author name in the invitation email or on the manuscript's title page; AH, high-prominence author name on manuscript only; AL, low-prominence author name on manuscript only; HH, high-prominence author name on both email and manuscript; LL, low-prominence author name on both email and manuscript.

<sup>a</sup>Fisher exact tests of acceptance rates across treatment conditions.

<sup>b</sup>Compared with LL,  $P = .32$ ; compared with HH,  $P = .003$ .

<sup>c</sup>Compared with HH,  $P = .001$ .

<sup>d</sup>Mann-Whitney tests of the distribution of recommendations (reject, major revision, minor revision, or accept) across treatment conditions.

<sup>e</sup>Compared with AA,  $P = .005$ .

<sup>f</sup>Compared with AL,  $P < .001$ ; compared with AA,  $P < .001$ .

<sup>g</sup>Economic Science Institute, Chapman University, Orange, CA, USA; <sup>h</sup>Department of Economics, University of Innsbruck, Innsbruck, Austria; <sup>i</sup>Institute of Banking and Finance, University of Graz, Graz, Austria

#### Conflict of Interest Disclosures

None reported.

**Funding/Support** Financial support from the Austrian Science Fund through SFB F63 is gratefully acknowledged.

**Role of the Funder/Sponsor** The funding organization did not play any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and the decision to submit the abstract for presentation.

**Additional Information** Jürgen Huber is a co-corresponding author.

#### Factor Analysis of Academic Reviewers' Ratings of Journal Articles on a 38-Item Scientific Quality Instrument

Guy Madison,<sup>1</sup> Erik J. Olsson<sup>2</sup>

**Objective** High scientific quality produces knowledge and is continuously assessed by the scientific community in multiple ways (eg, reviews, seminars, committees that control funding, publication of research results, and selection of individuals for awards, employment, and other positions). The quality of these assessments affects the quality of the research produced.<sup>1,2</sup> So how valid and reliable are they, and what weight do academics attach to different features of the research in their assessment of scientific quality?

**Design** A rating instrument was devised with 36 concrete and 2 overall quality indicators (**Table 12**), partly based on

the evaluation criteria in scientific journals' instructions for reviewers. Fifteen academics with a PhD in the social sciences were recruited through snowball sampling. They rated 60 randomly selected journal articles from the Swedish Gender Studies List of more than 12,000 publications about sex and gender across many disciplines (eg, gender studies [GS], sociology, history, medicine, and psychology) published between 2000 and 2010. Factor analysis was applied to identify the main dimensions that reviewers consider, and these were compared across research with different levels of GS perspective.<sup>3</sup> A high level of GS perspective was assigned to articles authored by those who self-identified as gender scholars or explicitly endorsed a gender perspective, a medium level to articles that reflected these values and beliefs, and a low level to the remaining articles. Statistical tests included Cronbach  $\alpha$  for interrater reliability,  $R^2$  for factor analysis, and Cohen  $d$  and  $P$  values for differences between GS groups.

**Results** Interrater reliability was high (Cronbach  $\alpha = 0.75\text{--}0.87$ ). Factor analysis suggested 3 dimensions: logic and clarity (eg, "there is a clear connection between theory and empirics, such that the data logically correspond to what is being claimed, hypothesized, or tested";  $R^2 = 0.44$ ); causality (eg, "a causal relation is mentioned or can be inferred, and the design and methods in general enable this to be evaluated";  $R^2 = 0.08$ ); and scope (eg, more data, units of analysis, generalizability, and better sampling;  $R^2 = 0.10$ ). The analysis could discriminate between 3 samples of 20 randomly selected articles from each of the populations with (1) a high level of GS perspective, (2) a medium level, and (3) no GS perspective on these 3 dimensions (logic and clarity,  $d = 0.62$ ,  $P < .001$ ; causality,  $d = 0.35$ ,  $P < .005$ ; and scope,  $d = 0.44$ ,  $P < .005$ ).

**Conclusions** The scientific quality instrument seems promising in terms of high reliability and convergent validity and can be used in any type of design (eg, experimental, cross-sectional, correlational, or descriptive) to assess overall and specific dimensions of scientific quality. Additional testing of the instrument is needed to compare different bodies of research, funding applications, project plans, and papers before and after peer review; assess group differences across topics, institutions, countries, and time; and compare mainstream vs controversial fields of research to further assess the role of research quality.

#### References

1. Finkel EJ, Eastwick PW, Reis HT. Replicability and other features of a high-quality science: toward a balanced and empirical approach. *J Pers Soc Psychol.* 2017;113(2):244-253. doi:10.1037/pspi0000075
2. Tuffaha HW, El Saifi N, Chambers SK, Scuffham PA. Directing research funds to the right research projects: a review of criteria used by research organisations in Australia in prioritising health research projects for funding. *BMJ Open.* 2018;8(12):e026207. doi:10.1136/bmjopen-2018-026207

**Table 12. Thirty-six Concrete and 2 Overall Quality Indicators**

1. What is the scope of this study in terms of being applicable or relevant to people in general? Consider the group or type of people to which the knowledge may apply, given that it were reliable. Note that the first option is qualitatively different from the others; it refers to the group for which empirical data are provided or presented, regardless of its size.	21. The aim(s) of the study were achieved.
2. What is the scope of this study in terms of being relevant at different times in the past?	22. Positive consistency of causal relations. A causal relation is mentioned or can be inferred, and the design and methods in general enable it to be evaluated.
3. What is the scope of this study in terms of being applicable or relevant at different times into the future?	23. Negative consistency of causal relations. The aims of the study could feasibly have been addressed by testing a causal relation, but no causal relation is mentioned or can be inferred.
4. What is the scope of this study in terms of being applicable or relevant to different geographical regions?	24. The study demonstrates a high level of reflexivity.
5. There is a clear connection between theory and empirics, such that the data logically correspond to what is being claimed, hypothesized, or tested. Tick 1 if either theory or empirics is wanting.	25. If you were asked to review this paper for a respectable scientific journal appropriate for the topic, how would you evaluate it?
6. It would be possible to test the claims, hypotheses, or theoretical assumptions made in this paper, regardless of whether testing was actually done.	26. The presentation, methods, analysis, and conclusions are objective, balanced, and unbiased and appear not to reflect the particular interest of any group or individual.
7. The claims, hypotheses, or theoretical assumptions made in this paper were tested by the study being described.	27. All claims, analyses, and conclusions are critically examined, and the potential weaknesses and limitations of the present study have been appropriately considered.
8. The writing is clear, accessible, and to the point.	28. The results are of such a character that they make meta-analysis possible. Note that meta-analysis is not limited to interventions, but to any outcome; see instructions for more details.
9. Concepts and constructs are clear and well defined.	29. There are unjustified claims, explicit or implicit, regarding generalizability.
10. The study contributes something new, whether a new problem, a new approach, new data, a new theory, or a new explanation.	30. The study is conducted and reported such that it can be replicated (regardless of the outcome of a possible replication).
11. The aim(s) of the study are clearly presented.	31. There are unjustified claims, explicit or implicit, regarding causality (ie, the method, design, or data do not allow for it).
12. The design of the study is appropriate to address its aims (by <i>design</i> is meant the whole use of data, manipulations, logic, and methods in general to address correlational, causal, or other relationships).	32. The reliability of the study is high: it demonstrates awareness of the issue of reliability and an effort to maximize it.
13. The paper demonstrates adequate knowledge of previous research and theory and builds on this in its argumentation and conclusions.	33. The validity of the study is high: it demonstrates awareness of the issue of validity and an effort to maximize it.
14. Citations of other work are appropriate, motivated, and sufficient.	34. The design of the study. If several types apply, choose the one that is most critical for the central claims or findings.
15. The data analysis is relevant, suitable, and clearly described.	35. Origin of the data in the study.
16. It is possible to obtain valid answers to the questions posed by means of the applied design and methods in general.	36. Type of data in the study.
17. The paper is logically organized, easy to follow, and easy to search for particular information.	37. What is the size of the population to which one can generalize the findings? Note that the first option is qualitatively different from the others; it refers to the group for which empirical data are provided or presented, regardless of its size.
18. The information and the statements in the paper are accurate.	38. Overall impression.
19. The argumentation is clear, rigorous, coherent, consistent, and logical.	Other questions include: How much time did you spend on assessing this text? How much of the text did you read?
20. Is the length of the text appropriate for the message conveyed?	

3. Söderlund T, Madison G. Objectivity and realms of explanation in academic journal articles concerning sex/gender: a comparison of gender studies and the other social sciences. *Scientometrics*. 2017;112(2):1093-1109.  
doi:10.1007/s11192-017-2407-x

<sup>1</sup>Department of Psychology, Umeå University, Umeå, Sweden, guy.madison@umu.se; <sup>2</sup>Department of Philosophy, Lund University, Lund, Sweden

**Conflict of Interest Disclosures** None reported.

### A Synthesis of Studies on Changes Manuscripts Underwent Between Submission or Preprint Posting and Peer-Reviewed Journal Publication

Mario Malički,<sup>1</sup> Ana Jerončić,<sup>2</sup> Gerben ter Riet,<sup>3,4</sup> Lex Bouter,<sup>5,6</sup> John P. A. Ioannidis,<sup>1,7,8,9,10</sup> IJsbrand Jan Aalbersberg,<sup>11</sup> Steven N. Goodman<sup>1,7,8</sup>

**Objective** The ability of peer review to improve the scientific endeavor (eg, the conduct, reporting, and validity of study

findings) has been questioned,<sup>1</sup> and calls have been made to showcase changes that occurred to each study due to peer review.<sup>2</sup> Until such transparency is achieved, identification and synthesis of studies that analyzed differences between preprints or manuscript versions submitted to journals and peer-reviewed publications is being undertaken.

**Design** In this stage of the living systematic review, studies were identified based on authors' knowledge of the field and by checking all research at peer review conferences (presented as podium presentations or posters in the European Union and USA). References and citations of identified studies were then checked. For all studies, the following was extracted: year of publication, sampling method, conflict of interest, funding, data and protocol sharing, number of analyzed version pairs, sample size calculation, scholarly discipline, method used to compare versions, variables (ie, manuscript sections) analyzed for changes, and metric with which the changes were quantified or qualitatively classified.

**Table 13. Main Findings and Characteristics of Studies That Analyzed Changes Between Submitted or Preprinted and Peer-Reviewed Journal Versions of Manuscripts<sup>a</sup>**

Study year	No. of analyzed version pairs	Discipline	Assessment method	Rating	Item with largest difference
1990	25	Health sciences	26-Item questionnaire	Yes, unclear, no	Addition of confidence intervals
1994	111	Health sciences	34-Item scale	5-Rating scale	Addition of limitations
1996	50	Health sciences	23-Item scale	5-Rating scale	Background section
2003	43	Health sciences	38-Item scale	5-Rating scale	Tone of conclusions
2007	115	Health sciences	36-Item scale	5-Rating scale	Study design reporting
2015	38	Social sciences	Manual comparison	Counting	Number of references
2019	38	Social sciences	Text similarity formulas	% of Content added	Theory and discussion session
2020	18	Health sciences	8-Item questionnaire	Yes, no, NA	Primary outcome reporting, and blinding

Abbreviation: NA, not applicable.

<sup>a</sup>This table presents findings from a sample of 8 of 25 studies analyzed to date.

**Results** Of 25 studies published from 1990 through the end of 2021, 16 (64%) analyzed changes between submitted and published papers and 9 (36%) between preprints and published papers. Most commonly, changes were analyzed by filling out questionnaires or scales separately for each of the 2 manuscript versions (11 [44%]) or by manual comparison of the 2 manuscript versions (6 [24%]). The median number of analyzed version pairs was 59 (IQR, 41-122). Most studies analyzed changes that occurred in health (18 [72%]) or social sciences (4 [16%]) manuscripts. Overall, studies' conclusions indicated very high similarity between version pairs, with the largest changes occurring in introduction and discussion sections. Examples of items for which most changes were found are presented in **Table 13**.

**Conclusions** The current results indicate that submitted or preprinted manuscript versions and their peer-reviewed journal version are very similar, with main (analysis) methods and main findings rarely changing. Quantification of these results is pending. Large differences between studies, type of manuscript changes, and methods with which they were measured indicate greater need for collaboration in the peer review field and creation of the core outcomes measures for manuscript version changes.

## References

1. Tennant JP, Ross-Hellauer T. The limitations to our understanding of peer review. *Res Integr Peer Rev*. 2020;5(1):6. doi:10.1186/s41073-020-00092-1
2. Limbu S. Building trust in peer review: a Q&A with Dr Mario Malički. BioMed Central. September 18, 2020.

Accessed June 24, 2022. <http://blogs.biomedcentral.com/on-medicine/2020/09/18/building-trust-in-peer-review-a-q-a-with-dr-mario-malicki/>

<sup>1</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA, mario.malicki@mefst.hr; <sup>2</sup>Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia; <sup>3</sup>Urban Vitality Centre of Expertise, Amsterdam University of Applied Sciences, Amsterdam, the Netherlands; <sup>4</sup>Amsterdam University Medical Centers, Department of Cardiology, Amsterdam, the Netherlands; <sup>5</sup>Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, the Netherlands; <sup>6</sup>Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, the Netherlands; <sup>7</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA;

<sup>8</sup>Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA; <sup>9</sup>Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA; <sup>10</sup>Department of Statistics, Stanford University School of Humanites and Sciences, Stanford, CA, USA; <sup>11</sup>Elsevier, Amsterdam, the Netherlands

**Conflict of Interest Disclosures** IJsbrand Jan Aalbersberg is senior vice president of research integrity at Elsevier. Mario Malički is a co-editor in chief of *Research Integrity and Peer Review*. Lex Bouter, John P. A. Ioannidis, and Steven N. Goodman are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract.

**Funding/Support** Elsevier funding was awarded to Stanford University for a METRICS postdoctoral position that supported Mario Malički's work on the project.

**Role of the Funder/Sponsor** IJsbrand Jan Aalbersberg is an employee of Elsevier and had a role in the design and conduct of the study; management and interpretation of the data; review and approval of the abstract; and decision to submit the abstract for presentation.

## Peer Reviewed Evaluation of Registered End Points of Randomized Trials (the PRE-REPORT Study)

Christopher W. Jones,<sup>1</sup> Amanda C. Adams,<sup>2</sup> Benjamin S. Misemer,<sup>3</sup> Mark A. Weaver,<sup>4</sup> Sara Schroter,<sup>5</sup> Hayat Khan,<sup>1</sup> Benyamin Margolis,<sup>6</sup> David L. Schriger,<sup>7</sup> Timothy F. Platts-Mills<sup>8</sup>

**Objective** To test whether providing information from relevant clinical trial registry entries to peer reviewers evaluating trial manuscripts decreases discrepancies between registered and published trial outcomes.<sup>1,2</sup>

**Design** This stepped-wedge, cluster-randomized trial studied clusters comprised of eligible clinical trial manuscripts submitted to each of 13 high-impact medical journals across a range of medical specialties between November 1, 2018, and October 31, 2019.<sup>3</sup> All participating journals routinely published results from clinical trials. Each journal began the study in the control phase; journals were crossed over into the intervention phase in random order. Manuscripts submitted during the study period were included if they presented results from the primary analysis of a clinical trial and underwent external peer review. Peer reviewers were not told that a study was taking place. Preexisting peer review

practices were unchanged during the control phase. Peer reviewers for intervention-phase manuscripts received a data sheet describing whether the trial was registered, the initial registration and enrollment dates, and the registered primary outcome(s) when enrollment began. Decision editors had access to the registry data sheets. The primary end point was the presence of a published primary outcome consistent with a prospectively defined primary outcome in the study's trial registry (ie, registered before enrollment began), as determined independently by 2 outcome assessors blinded to each manuscript's study arm. Linear mixed models were used to estimate outcome differences between intervention- and control-condition manuscripts. For the primary end point, use of a 1-sided test at the 5% level was prespecified, with corresponding 90% CIs, based on the assumption that the intervention was unlikely to increase outcome inconsistencies.

**Results** The study included 419 submitted manuscripts. Participating journals published 105 of 243 control-phase manuscripts (43%) and 68 of 176 intervention-phase manuscripts (39%) (model-estimated difference between intervention and control, -10%; 95% CI, -25% to 4%). Among the 173 accepted manuscripts, published primary outcomes were consistent with clearly defined, prospectively registered primary outcomes in 40 of 105 control-phase manuscripts (38%) and 27 of 68 intervention-phase manuscripts (40%). There was no statistically significant difference between intervention and control phases in the primary end point (estimated difference, -6%; 90% CI, -27% to 15%; 1-sided  $P = .68$ ). Prospectively registered trials were more likely to be published (117 of 251 trials [47%]) than unregistered trials (7 of 33 trials [21%]) (model-estimated difference, 29%; 95% CI, 10%-47%), but no significant difference was observed between prospectively and retrospectively registered trials (49 of 135 trials [36%] accepted; model-estimated difference, 9%; 95% CI, 1%-20%).

**Conclusions** The results do not support provision of a data sheet with clinical trial registration details during peer review to increase agreement between prospectively registered and published trial outcomes. The high prevalence of retrospective registration and discrepancies between registered and published trial outcomes necessitates identification of effective interventions for these problems.

## References

- Trinquart L, Dunn AG, Bourgeois FT. Registration of published randomized trials: a systematic review and meta-analysis. *BMC Med*. 2018;16(1):173. doi:10.1186/s12916-018-1168-6
- Wager E, Williams P, "Hardly worth the effort"? medical journals' policies and their editors' and publishers' views on trial registration and publication bias: quantitative and qualitative study. *BMJ*. 2013;347:f5248. doi:10.1136/bmj.f5248

3. Ellenberg SS. The stepped-wedge clinical trial: evaluation by rolling deployment. *JAMA*. 2018;319(6):607-608. doi:10.1001/jama.2017.21993

<sup>1</sup>Department of Emergency Medicine, Cooper Medical School of Rowan University, Camden, NJ, USA, jones-christopher@cooperhealth.edu; <sup>2</sup>Medical Library, Cooper Medical School of Rowan University, Camden, NJ, USA; <sup>3</sup>Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, USA; <sup>4</sup>Department of Mathematics and Statistics, Elon University, Elon, NC, USA; <sup>5</sup>The BMJ, BMA House, London, UK; <sup>6</sup>Rollins School of Public Health, Emory University, Atlanta, GA, USA; <sup>7</sup>Department of Emergency Medicine, University of California Los Angeles, Los Angeles, CA, USA; <sup>8</sup>Ophirex, Inc, Corte Madera, CA, USA

**Conflict of Interest Disclosures** Christopher W. Jones reported receiving grants from AstraZeneca, Abbott, Vapotherm, and Ophirex outside the submitted work. Sara Schroter is a full-time employee at *The BMJ*. Benjamin Margolis is an employee of the Department of Health and Human Services in the Office of Research Integrity. David L. Schriger is an associate editor at *JAMA* and a deputy editor at *Annals of Emergency Medicine*. Timothy F. Platt-Mills is an employee of Ophirex.

**Funding/Support** This work was supported by the US Department of Health and Human Services Office of Research Integrity (grant ORIIR180039).

**Role of the Funder/Sponsor** The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the abstract for presentation.

**Acknowledgment** We thank the editors and staff members of the following journals for their participation in this study: *Academic Emergency Medicine*, *American Journal of Transplantation*, *Annals of Emergency Medicine*, *Archives of Physical Medicine and Rehabilitation*, *British Journal of Ophthalmology*, *Clinical Orthopaedics and Related Research*, *Gastroenterology*, *Heart*, *International Journal of Cancer*, *Journal of the American College of Surgeons*, *Neurology*, *Surgery*, and *Thorax*.

**Additional Information** The contents of this manuscript are those of the authors and do not represent the official views of, nor an endorsement by, the Office of the Assistant Secretary of Health, the Department of Health and Human Services, or the US government. The study was registered at ISRCTN41225307.

## Dissemination of Clinical Trial Findings

### Analysis of Reporting Consistency Between Clinical Trials Presented at Major Medical Conferences, Their Corresponding Publications, and Press Releases

Anisa Rowhani-Farid,<sup>1</sup> Kyungwan Hong,<sup>1</sup> Mikas Grewal,<sup>2</sup> Jesse Reynolds,<sup>3</sup> Audrey D. Zhang,<sup>4</sup> Joshua D. Wallach,<sup>5</sup> Joseph S. Ross<sup>7,8,9</sup>

**Objective** Clinical trial integrity is compromised when investigators selectively report or misreport results, which leads to inaccurate claims of benefit and/or extrapolations from incomplete data.<sup>1-3</sup> This study examined the extent to which trials presented at major international medical conferences in 2016 consistently reported their study design,

end points, and results across conference abstracts, published article abstracts, and press releases.

**Design** A cross-sectional analysis of trials presented at 12 conferences in the US in 2016 was conducted. Conferences were identified from a list prepared by the Healthcare Convention and Exhibitors Association and were included if abstracts were publicly reported. From these conferences, all late-breaker trials were included and other trials were randomly selected, bringing the total sample to 25 abstracts per conference. First, it was determined whether trials were registered and reported results in an International Committee of Medical Journal Editors–approved trial registry. Second, it was determined whether trial results were published in a peer-reviewed journal. Finally, information on trial media coverage and press releases was collected using LexisNexis. For all published trials, the consistency of reporting of the following characteristics was examined, through comparison of the trials' conference and publication abstracts: primary efficacy end point, safety end point, sample size, follow-up period, primary end point effect size, and characterization of results (comparisons were made for this characteristic across press releases too, if any). Authors determined consistency of reporting when identical information was presented across abstracts (and press releases). Primary analyses were descriptive; secondary analyses included  $\chi^2$  tests and multiple logistic regression.

**Results** The sample comprised 240 trials presented at 12 conferences. Of these, 208 trials (86.7%) were registered, 95 (39.6%) reported summary results in a registry, and 177 (73.8%) were published; 82 trials (34.2%) were covered by the media, and 68 (28.3%) had press releases. Among the 177 published trials (**Table 14**), 171 (96.6%) reported consistent primary efficacy end points across abstracts, whereas 96 of 128 trials (75.0%) reported consistent safety outcomes. There were 107 of 172 trials (62.2%) with consistent sample sizes across abstracts, 101 of 137 trials (73.7%) that reported their follow-up periods consistently, 92 of 175 trials (52.6%) that described their effect sizes consistently, and 157 of 175 trials (89.7%) that characterized their results consistently. Among the trials that were published and had press releases, 32 of 32 (100%) characterized their results consistently across conference and publication abstracts and press releases. No trial characteristics were associated with reporting primary efficacy end points consistently.

**Conclusions** This study demonstrates that trials are consistently reporting primary efficacy end points and results characterization. Lower consistency rates for other characteristics indicate that trial presentations are less likely to report safety end points and that authors could be presenting preliminary data at conferences with shorter follow-up periods and smaller sample sizes, owing to incomplete patient recruitment, and consequently, varying effect sizes. However, this does not rule out the possibility of misreporting at conferences or publications.

**Table 14. Consistency of Reporting Between Conference and Publication Abstracts for the Clinical Trials Included in the Study Sample**

Characteristic	Consistency, No./total No. (%)
Primary efficacy end point	171/177 (96.6) <sup>a</sup>
Safety end point	96/128 (75.0) <sup>b</sup>
Sample size	107/172 (62.2) <sup>c</sup>
Follow-up period	101/137 (73.7) <sup>d</sup>
Effect size	92/175 (52.6) <sup>e</sup>
Results characterization	157/175 (89.7) <sup>f</sup>

Explanation of why denominators differ for each row:  
<sup>a</sup>Among 240 trials, 177 were published.  
<sup>b</sup>Among 177 published trials, 128 reported safety end points in either the conference abstract or the publication abstract or across both locations.  
<sup>c</sup>Among 177 published trials, 172 reported sample sizes across both conference abstracts and publication abstracts.  
<sup>d</sup>Among 177 published trials, 137 reported follow-up periods across both conference abstracts and publication abstracts.  
<sup>e</sup>Among 177 published trials, 175 reported effect sizes across both conference abstracts and publication abstracts.  
<sup>f</sup>Among 177 published trials, 175 characterized results across both conference abstracts and publication abstracts.

## References

1. Boutron I. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303:2058-2064. doi:10.1001/jama.2010.651
2. Chiu K, Grundy Q, Bero L. “Spin” in published biomedical literature: a methodological systematic review. *PLoS Biol*. 2017;15:e2002173. doi:10.1371/journal.pbio.2002173
3. Diong J, Butler AA, Gandevia SC, Héroux ME. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. *PLoS One*. 2018;13:e0202121. doi:10.1371/journal.pone.0202121

<sup>a</sup>Restoring Invisible and Abandoned Trials Support Center, Department of Pharmaceutical Health Services Research, University of Maryland, Baltimore, MD, USA, arowhani@rx.umaryland.edu; <sup>b</sup>Yale School of Medicine, New Haven, CT, USA; <sup>c</sup>Yale School of Public Health, New Haven, CT, USA; <sup>d</sup>Department of Internal Medicine, Duke University School of Medicine, Durham, NC, USA; <sup>e</sup>Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA; <sup>f</sup>Section of General Internal Medicine, Yale School of Medicine, New Haven, CT, USA; <sup>g</sup>National Clinician Scholars Program, Yale School of Medicine, New Haven, CT, USA; <sup>h</sup>Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA; <sup>i</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA

**Conflict of Interest Disclosures** The Laura and John Arnold Foundation funds the Restoring Invisible & Abandoned Trials Support Center, which supports the salaries of Anisa Rowhani-Farid and Kyungwan Hong. Kyungwan Hong was supported by the Food and Drug Administration (FDA) of the US Department of Health and Human Services (HHS) as part of a financial assistance award U01FD005946, unrelated to this manuscript, totaling \$5000 with 100% funded by the FDA/HHS. Audrey D. Zhang currently receives research support from the National Institute on Aging through the Duke Creating Alzheimer's Disease and Related Dementias Researchers for the Next Generation—Stimulating Access to Research in Residency (CARiNG-StARR) program (R38AG065762). Joshua D. Wallach is supported by the FDA and the National Institute on Alcohol Abuse and Alcoholism of the National

Institutes of Health (NIH) under award 1K01AA028258. Joseph S. Ross currently receives research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology, from the FDA for the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation program (U01FD005938), from the Agency for Healthcare Research and Quality (R01HS022882), from the National Heart, Lung, and Blood Institute of the NIH (R01HS025164, R01HL144644), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International; in addition, Joseph S. Ross is an expert witness at the request of relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. No other disclosures were reported.

**Additional information** Kyungwan Hong is a co-corresponding author.

### Evaluating Prospective Study Registration and Result Reporting of Trials Conducted in Canada From 2009-2019

Mohsen Alayche,<sup>1,2</sup> Kelly D. Cobey,<sup>3,4</sup> Jeremy Y. Ng,<sup>1</sup> Clare L. Ardern,<sup>5,6</sup> Karim M. Khan,<sup>7</sup> An-Wen Chan,<sup>8,9</sup> Ryan Chow,<sup>1,2</sup> Mouayad Masalkhi,<sup>10</sup> Ana Patricia Ayala,<sup>11</sup> Sanam Ebrahimzadeh,<sup>1</sup> Jason Ghossein,<sup>2</sup> Ibrahim Alayche,<sup>2</sup> Jessie V. Willis,<sup>1,2</sup> David Moher<sup>1,4</sup>

**Objective** The objective of this study was to determine the proportion of clinical trials conducted in Canada that adhered to the World Health Organization's registration and reporting best practices. The specific contributing factors that impaired adherence to those best practices were highlighted.

**Design** All registered clinical trials on ClinicalTrials.gov conducted in Canada as of 2009 and completed by 2019 were identified. A cross-sectional analysis of those trials measured prospective registration (as opposed to retrospective registration), subsequent result reporting in the registry at any point, and subsequent publication of study findings at any point. Unregistered and/or incomplete clinical trials were excluded. This means that the results likely underestimate the true prevalence of nonreporting of trials conducted in Canada. The lead sponsor, phase of study, countries involved, total patient enrollment, number of arms, type of masking, type of allocation, year of completion, and patient demographics were examined as potential effect modifiers to these best practices.

**Results** A total of 6720 trials met the inclusion criteria. From 2009 to 2019, 59% (n = 3967) of clinical trials were registered prospectively and 39% (n = 2642) reported their results in the registry. Of the trials registered between 2009 and 2014, 55% (n = 1482) were subsequently published in an academic journal. Over time, the annual rate of compliance with study registration and subsequent publication of findings improved (increased). However, there was a downward trend over time in results being reported in the registry. Of the 3763 trials conducted exclusively in Canada, 3% (n = 123) met all 3 criteria of prospective registration, reporting in the registry, and publishing findings. In contrast, of the remaining 2957

trials with both Canadian and international sites, 41% (n = 1238) had an overall compliance to these 3 criteria.

**Conclusions** Canadian clinical trials substantially lacked adherence to study registration and reporting best practices. Knowledge of this widespread noncompliance should motivate stakeholders in the Canadian clinical trial ecosystem to address and continue to monitor this problem. The data presented provide a baseline against which to compare any improvement in the registration and reporting of clinical trials in Canada.

<sup>1</sup>Centre for Journalalogy, Ottawa Methods Centre, Ottawa Hospital Research Institute, Ottawa, ON, Canada, mohsen.alayche@gmail.com; <sup>2</sup>Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>3</sup>University of Ottawa Heart Institute, Ottawa, ON, Canada; <sup>4</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>5</sup>Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Australia; <sup>6</sup>Department of Family Practice, University of British Columbia, Vancouver, BC, Canada; <sup>7</sup>Department of Family Practice and School of Kinesiology, University of British Columbia, Vancouver, BC, Canada; <sup>8</sup>Department of Medicine, Women's College Research Institute, Toronto, ON, Canada; <sup>9</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; <sup>10</sup>School of Medicine, University College Dublin, Dublin, Ireland; <sup>11</sup>Gerstein Science Information Centre, University of Toronto, Toronto, ON, Canada

**Conflict of Interest Disclosures** David Moher is an associate director and member of the Peer Review Congress advisory board and An-Wen Chan is a member of the Peer Review Congress advisory board but neither were not involved in the review or decision for this abstract. No other conflicts of interest to declare.

**Funding/Support** This study was funded by the University of Ottawa Summer Studentship Program. This program provides a stipend for a University of Ottawa medical student's summer research project.

**Role of the Funder/Sponsor** The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

### Dissemination of the Results of Pediatric Clinical Trials Funded by the US National Institutes of Health

Chris A. Rees,<sup>1</sup> Adrianna Westbrook,<sup>2</sup> Florence T. Bourgeois<sup>3</sup>

**Objective** The National Institutes of Health (NIH) is the largest source of government funding for biomedical research in the world, although only a small proportion of its funding is dedicated to pediatric research. Timely dissemination of results for NIH-funded pediatric clinical trials is imperative to uphold the integrity of scientific evidence, maintain ethical obligations to trial participants, and ensure evidence-based clinical care. Since 2017, the NIH has required NIH-funded clinical trials to be registered and summary results submitted to ClinicalTrials.gov, generally not later than 1 year after primary trial completion.<sup>1</sup> In this study, the dissemination of NIH-funded pediatric clinical trial findings was investigated through results submitted to ClinicalTrials.gov and in publications in peer-reviewed medical journals.

**Table 15. Publication of National Institutes of Health–Funded Pediatric Clinical Trials**

Grant and trial characteristics	No. (%)			Adjusted odds ratio for nonpublication (95% CI)	P value
	All trials (N = 421)	Published trials (n = 206)	Unpublished trials (n = 215)		
Year of grant completion					
2017	142 (33.7)	81 (39.3)	61 (28.4)	1 [Reference]	
2018	139 (33.0)	75 (36.4)	64 (29.8)	1.22 (0.70-2.13)	.48
2019	140 (33.3)	50 (24.3)	90 (41.9)	3.52 (2.02-6.15)	<.001
Funding type					
R01 awards	205 (48.7)	115 (55.8)	90 (41.9)	1 [Reference]	
K awards	47 (11.2)	15 (7.3)	32 (14.9)	1.31 (0.44-3.84)	.63
Other awards	169 (40.1)	76 (36.9)	93 (43.3)	1.05 (0.51-2.17)	.90
Funding amount, \$ (in millions)					
0 to 1	176 (41.8)	69 (33.5)	107 (49.8)	1 [Reference]	
>1 to 3	129 (30.6)	65 (31.6)	64 (29.8)	0.91 (0.39-2.17)	.84
>3	116 (27.6)	72 (35.0)	44 (20.5)	0.71 (0.31-1.63)	.42
Intervention type					
Behavioral	321 (76.3)	154 (74.8)	167 (77.7)	1.25 (0.69-2.27)	.46
Nonbehavioral <sup>a</sup>	100 (23.8)	52 (25.2)	48 (22.3)	1 [Reference]	
Study design <sup>b</sup>					
Randomized clinical trial	311 (75.3)	147 (71.4)	164 (79.2)	1.52 (0.74-3.15)	.26
Cluster randomized trial	54 (13.1)	35 (17.0)	19 (9.2)	0.85 (0.31-2.32)	.75
Nonrandomized trial	48 (11.6)	24 (11.7)	24 (11.6)	1 [Reference]	
Trial enrollment, No. of participants <sup>c</sup>					
<100	135 (36.1)	61 (29.8)	74 (43.8)	1 [Reference]	
100 to 299	117 (31.3)	67 (32.7)	50 (29.6)	0.66 (0.36-1.21)	.18
>300	122 (32.6)	77 (37.6)	45 (26.6)	0.69 (0.35-1.38)	.30
Youngest participant age group <sup>d</sup>					
Neonate (aged 0-30 d)	84 (20.5)	46 (22.4)	38 (18.6)	1 [Reference]	
Infants and children (aged 31 d to 11 y)	213 (52.1)	106 (51.7)	107 (52.5)	1.18 (0.65-2.13)	.60
Adolescent (aged 12-17 y)	112 (27.4)	53 (25.9)	59 (28.9)	1.21 (0.60-2.42)	.59
Trial end point <sup>e</sup>					
Clinical event	157 (38.2)	97 (47.1)	60 (29.3)	1 [Reference]	
Surrogate measure	184 (44.8)	73 (35.4)	111 (54.2)	2.09 (1.26-3.45)	.004
Scale	70 (17.0)	36 (17.5)	34 (16.6)	1.31 (0.69-2.47)	.41

<sup>a</sup>Includes drugs, devices, procedures, and dietary interventions.<sup>b</sup>Excluding 8 grants with undeclared study designs.<sup>c</sup>Limited to grants with publications and/or ClinicalTrials.gov registration (n = 374).<sup>d</sup>Excluding 12 grants that did not report specific age groups.<sup>e</sup>Excluding 10 grants with unknown end points.

**Design** A cross-sectional analysis of NIH grants funding pediatric clinical trials with funding completed from January 1, 2017, to December 31, 2019, was conducted. Grants funding pediatric trials in NIH RePORTER and the status of registration and results submission in ClinicalTrials.gov were determined. Publications were identified in PubMed as of February 28, 2022. Time to results reporting in ClinicalTrials.gov at 12 months and 24 months since primary trial completion, and factors associated with nonpublication from an a priori set of variables using multivariable logistic regression were determined.

**Results** Among 3408 pediatric grants completed during the study period, 421 (12.4%) supported an interventional study (**Table 15**). Mental and behavioral health conditions (70 trials [16.6%]), obesity (56 [13.3%]), and substance use (40 [9.5%]) were the most commonly studied conditions. Greater than three-quarters of the trials studied behavioral interventions (321 [76.2%]). There were 360 trials (85.5%) registered in ClinicalTrials.gov, of which 229 (63.6%) were registered prospectively (ie, within 21 days of study start). Results were submitted for 16.6% of trials (95% CI, 14.4%-18.8%) by 12 months and 22.8% (95% CI, 20.2%-25.5%) by

24 months. Publications were available for 11.8% of trials (95% CI, 10.2%-13.5%) at 12 months and 27.7% (95% CI, 25.3%-30.1%) at 24 months. Overall, 56,833 participants were enrolled in trials that remained unpublished after a median follow-up of 33 months since trial completion. Trials were more likely to remain unpublished if the trial end point was a surrogate measure (adjusted odds ratio, 2.09; 95% CI, 1.26-3.45) compared with a clinical event.

**Conclusions** Despite policies promoting dissemination of NIH-funded clinical trial results, less than a third of pediatric trials reported results in ClinicalTrials.gov or published findings in peer-reviewed journals 2 years after completion. To maximize the impact of pediatric clinical trials, additional efforts are needed to improve reporting practices and advance translation of research findings into evidence-based clinical care. Such efforts may include the withholding of additional funding unless trials have been prospectively registered and results publicly reported in a timely manner.

## Reference

1. NIH policy on the dissemination of NIH-funded clinical trial information. National Institutes of Health. Updated July 24, 2017. Accessed March 5, 2022. <https://grants.nih.gov/policy/clinical-trials/reporting/understanding/nih-policy.htm>

<sup>1</sup>Division of Pediatric Emergency Medicine, Emory University School of Medicine, Children's Healthcare of Atlanta, Atlanta, GA, USA, chris.rees@emory.edu; <sup>2</sup>Pediatric Biostatistics Core, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA; <sup>3</sup>Pediatric Therapeutics and Regulatory Science Initiative, Computational Health Informatics Program, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

**Conflict of Interest Disclosures** No authors have potential conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject of this abstract.

## Grant Review and Funded Research

### A Bayesian Approach to Reduce Bias in the Ranking of Peer-Reviewed Grant Proposals Submitted to the Swiss National Science Foundation

Rachel Heyard,<sup>1,2</sup> Manuela Ott,<sup>2</sup> Janine Bühler,<sup>2</sup> Georgia Salanti,<sup>3</sup> Matthias Egger<sup>2,3</sup>

**Objective** Funding agencies rely on peer review to select grant proposals deserving of funding. Peer review has limitations and may be biased. The interrater reliability between reviewers and panels is low, particularly for proposals near the funding line. A method for ranking grant proposals was developed to reduce bias and accommodate the uncertainty in evaluation and funding decisions.

**Design** Proposals submitted to a call were peer reviewed and later discussed in evaluation panels. After the discussion, all panel members electronically rated each proposal on a scale from 1 to 9. Panel members with conflicts of interest did not participate in the discussion or vote. Simple averages can

introduce bias if panel members with distinct behaviors (eg, very critical or very generous) cannot vote. Furthermore, there may be uncertainty in ranking. A ranking approach based on a flexible bayesian hierarchical model (BHM) was developed to avoid the long and biased discussions of proposals around the funding line and was compared with standard procedure. The BHM accounted for the correlated data structure due to the same panel members voting on a set of proposals and modeled explicitly the uncertainty present at different stages of the evaluation process. As such, the BHM described the whole distribution of the rank of each proposal. The 50% credible intervals around the ranks helped assign the proposals into 3 groups: accepted, random selection, and rejected. The random selection group was composed of proposals of similar quality near the funding line. The approach was flexible and could take special cases into account. For example, if proposals were discussed in subpanels, the further level of dependency could be accounted for in the model. The convergence of the BHM was investigated using Gelman-Rubin convergence diagnostics.

**Results** The approach was simulated in the Swiss National Science Foundation (SNSF) early career fellowship scheme call of February 2020. A total of 181 fellowship grants were submitted to 5 disciplinary panels—humanities (23 fellowship grants), social sciences (38), medicine (35), biology (35), and STEM (50); 79 were discussed in the panel, and 32 were funded (**Figure 12**). A funding line was drawn based on the available budget. A random selection group was identified for 2 panels: medicine and STEM. Applications to other funding schemes will be presented.

**Conclusions** In this study, a method to address the limitations of peer review of good but not outstanding proposals was developed. The bayesian ranking approach ensured a transparent translation of votes into a ranking of the proposals. The approach was extensively discussed with stakeholders in 2021, and the Research Council of the SNSF recently decided to adopt it in most of its funding schemes later in 2022.

<sup>1</sup>Center for Reproducible Science, University of Zurich, Zurich, Switzerland, rachel.heyard@uzh.ch; <sup>2</sup>Swiss National Science Foundation, Bern, Switzerland; <sup>3</sup>Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

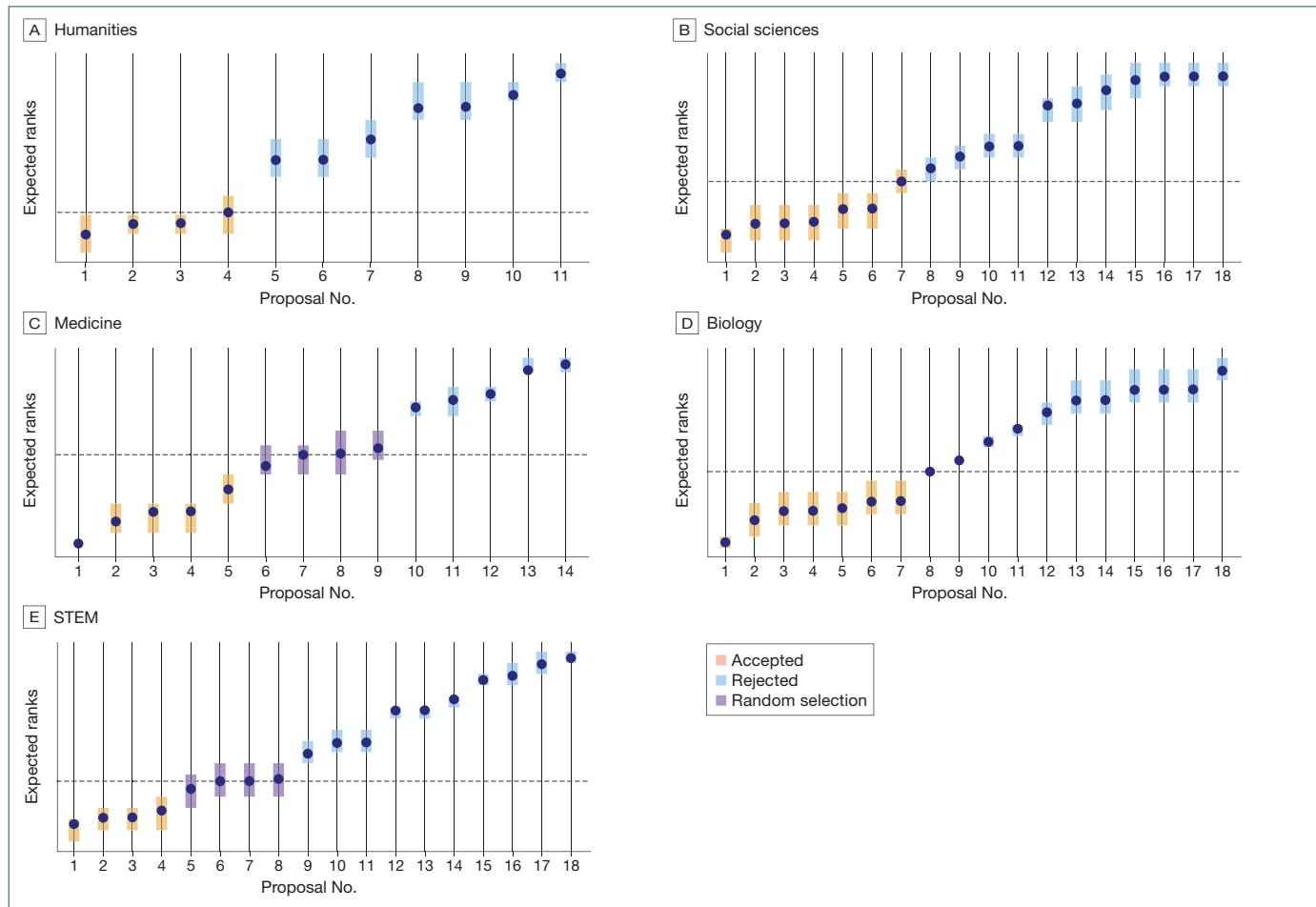
**Conflict of Interest Disclosures** Matthias Egger is the president of the Research Council of the Swiss National Science Foundation. No other disclosures were reported.

### Comparison of Availability of Trial Results in ClinicalTrials.gov and PubMed by Funder Type and Trial Primary Completion Date

Julianne T. Nelson,<sup>1</sup> Tony Tse,<sup>2</sup> Yvonne Puplampu-Dove,<sup>1</sup> Elisa Golfinopoulos,<sup>1</sup> Deborah A. Zarlin<sup>3</sup>

**Objective** The proportion and timing of results available for registered clinical trials was examined by data source (ie, ClinicalTrials.gov and PubMed) and funder type. Prior work assessed PubMed-indexed publications for National Institutes

**Figure 12. Bayesian Ranking Recommendations**



of Health (NIH)-funded trials completed by 2008<sup>1</sup>; published trial results, by academic and industry funder, with results posted on ClinicalTrials.gov in 2012<sup>2</sup>; and industry- and nonindustry-funded trials completed by 2015 with published or posted results.<sup>3</sup> This study updates prior work and appears to be the first analysis stratified by 4 funder types: NIH, non-NIH US federal agency, industry, and other (eg, foundation).

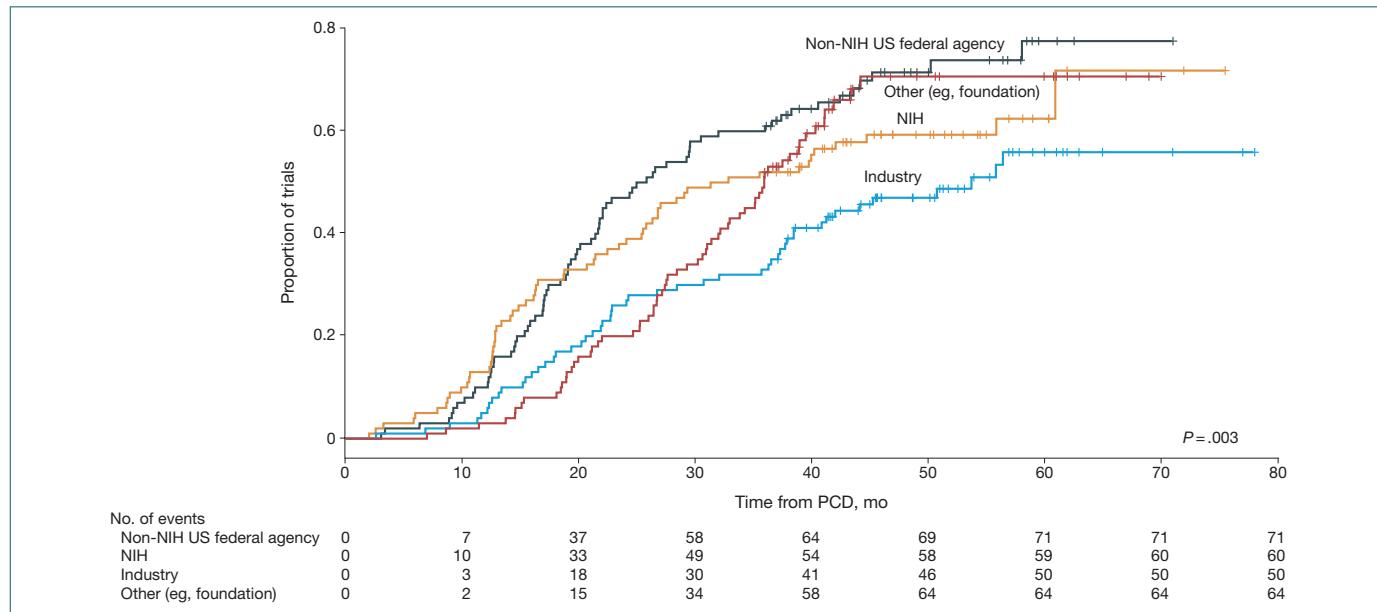
**Design** On August 1, 2021, a cross-sectional analysis was conducted of ClinicalTrials.gov-registered trials with at least 1 US facility, initiated on or after January 1, 2015, with primary completion dates (PCDs) up to August 1, 2018. A total of 100 trials from each key funder type was randomly sampled (**Figure 13**). To identify PubMed-indexed publications reporting primary outcomes between August 6 and October 8, 2021, 4 authors each reviewed registered information for 100 trials,<sup>1</sup> and the fifth resolved ambiguities. The median time for follow-up from PCD was 47.8 months (IQR, 41.9–56.8 months). The proportion of trials with results, by data source and funder type, was determined, and time from PCD to first results availability was evaluated using Kaplan-Meier analysis with R.

**Results** Results were identified for 245 of 400 trials (61.3%) at least 36 months after PCD (57 trials with results available from both ClinicalTrials.gov and PubMed), with 98 of 245

(40.0%) and 147 of 245 (60.0%) first available on ClinicalTrials.gov and PubMed, respectively. Among these 245 records, the median time from PCD to first results availability was 22.1 months (IQR, 14.9–32.9 months) from either data source, 18.8 months (IQR, 13.4–29.3 months) to posting on ClinicalTrials.gov (n = 117), and 25.6 months (IQR, 18.6–36.3 months) to PubMed-indexed publication (n = 185). The proportions of trials with results available by key funder type were 60 of 100 (NIH), 71 of 100 (non-NIH US federal agency), 50 of 100 (industry), and 64 of 100 (other/foundation). Time to first results availability by funder type showed significant differences ( $P = .003$ ) (**Figure 13**).

**Conclusions** The finding that 40.0% of sampled trials (98 of 245) with results were posted on ClinicalTrials.gov before publication suggests that searching both data sources maximizes discovery of trial results, which is generally consistent with prior findings.<sup>2</sup> Significant differences in proportion of results and time to first availability by key funder type suggest that sponsors and trialists may be affected by different factors based on funding source, such as policy and/or legal requirements. A better understanding of funding-specific factors could help improve overall results availability and timing. Larger sample sizes to validate these preliminary findings and research on the effectiveness of reporting requirements are needed. Finally, sampling trials conducted from January 2015 to August 2018 ensured recent

**Figure 13. Cumulative Proportion of 400 Trials With Results Available After PCD by Key Funder Type**



NIH indicates National Institutes of Health; PCD, primary completion date.

initiation and allowed time for publication but excluded longer-running trials.

## References

- Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross-sectional analysis. *BMJ*. 2012;344:d7292. doi:10.1136/bmj.d7292
- Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *PLoS Med*. 2013;10(12):e1001566. doi:10.1371/journal.pmed.1001566
- Zwierzyna M, Davies M, Hingorani AD, Hunter J. Clinical trial design and dissemination: comprehensive analysis of ClinicalTrials.gov and PubMed data since 2005. *BMJ*. 2018;361:k2130. doi:10.1136/bmj.k2130

<sup>1</sup>ICF, Bethesda, MD, USA, julianne.nelson@nih.gov; <sup>2</sup>The National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA; <sup>3</sup>Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard, Cambridge, MA, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This research was supported by the National Center for Biotechnology Information of the National Library of Medicine, National Institutes of Health.

**Role of the Funder/Sponsor** The National Library of Medicine, National Institutes of Health supported the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Additional Information** The views expressed in this abstract are those of the authors and do not necessarily reflect the views or policies of the National Institutes of Health. Tony Tse is a co-corresponding author.

## Funding Reporting Compliance in Metadata of Published Articles Supported by European and US Research Grants

Antonija Mijatović,<sup>1</sup> David G. Pina,<sup>2</sup> Ivan Buljan,<sup>1</sup> Ana Marušić<sup>1</sup>

**Objective** Inadequate compliance with reporting of findings in research articles may reduce the transparency of the published research and may negatively affect the ability of research funders to properly identify the body of knowledge associated with their grants. This investigation explored whether research funding was properly reported in publications by grant beneficiaries funded by the European Horizon (H2020) program (2014–2020) in comparison with those funded by the US National Institutes of Health (NIH) or the US National Science Foundation (NSF) during the same time period.

**Design** In this observational study, DOIs for publications from H2020 grants were identified using the official portal for European data. Data were collected from 2 funding programs, the Marie Skłodowska-Curie Actions (MSCA) and the European Research Council (ERC). The NIH grants were identified in PubMed, and NSF grants were found using the NSF Public Access Repository. Metadata were retrieved from Scopus using an application programming interface. The main outcome measure was the percentage of articles with correct textual funding declarations listed in the JavaScript Object Notation *fundText* metadata tags. Funding declarations were considered correct (accurate) if they contained the information on the funding agency and the correct grant number. For example, for ERC-funded articles, the funding declaration was deemed accurate if it contained (1) European Research Council or ERC, (2) Horizon 2020 or H2020, and (3) the correct grant number.

**Table 16. Accurate Reporting of Funding in Articles on Research Funded by the ERC, MSCA, NIH, and NSF**

Agency/year	Articles, No. (%)	Articles with accurately reported funding, No. (%) <sup>a</sup>
Total	209,635 (100.0)	120,051 (57.3)
Funding agency		
ERC	35,592 (17.0)	12,499 (35.1)
MSCA	22,325 (10.7)	11,327 (50.7)
NIH	85,933 (41.0)	45,153 (52.5)
NSF	65,785 (31.4)	51,072 (77.6)
Publication year		
2014	10 (0.0)	0
2015	595 (0.3)	36 (6.1)
2016	3807 (1.8)	610 (16.0)
2017	8735 (4.2)	2357 (27.0)
2018	13,366 (6.4)	5672 (42.4)
2019	16,880 (8.1)	7685 (45.5)
2020	90,492 (43.2)	55,863 (61.7)
2021	74,088 (35.3)	46,685 (63.0)
2022	1662 (0.8)	1143 (68.8)

Abbreviations: ERC, European Research Council; MSCA, Marie Skłodowska-Curie Action; NIH, US National Institutes of Health; NSF, US National Science Foundation.

<sup>a</sup>Defined as including the funding agency and the correct grant number in the metadata tags.

**Results** Of the 209,635 retrieved articles, 57.3% had accurately reported their funding (**Table 16**). The accuracy for NSF-funded articles was the highest, and that for MSCA and NIH-funded articles was higher than for ERC-funded articles ( $\chi^2 = 19,455.80$ ;  $P < .001$ ). The reporting accuracy of funding improved during the study years ( $\chi^2 = 10,553.71$ ;  $P < .001$ ). In logistic analysis, reporting accuracy was better for articles with a greater number of funding agencies (odds ratio [OR], 1.16; 95% CI, 1.16-1.17), smaller number of authors (OR, 0.94; 95% CI, 0.93-0.94), and later publication years (OR, 1.57; 95% CI, 1.55-1.58).

**Conclusions** Approximately half of the research articles supported by major funding agencies had adequate reporting of research funding. Research groups with fewer authors and those with more funding were better at adequate reporting. The accuracy of funding reporting improved over the years but was still suboptimal. Additional instructions and tools may be necessary to ensure that funding recipients properly acknowledge funding sources in their publications.

<sup>1</sup>Department of Research in Biomedicine and Health, Center for Evidence-Based Medicine, University of Split School of Medicine, Split, Croatia, antonija.mijatovic@mefst.hr;

<sup>a</sup>European Research Executive Agency, European Commission, Brussels, Belgium

**Conflict of Interest Disclosures** Ana Marušić occasionally serves as an expert for the European Research Executive Agency. Ana Marušić is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

**Funding/Support** This study was funded by the Croatian Science Foundation program, Professionalism in Health—Decision-making in Practice and Research (No. IP-2019-04-4882).

**Role of the Funder/Sponsor** The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Additional Information** All views expressed in this abstract are strictly those of the authors and may in no circumstances be regarded as an official position of the European Research Executive Agency or the European Commission.

## Saturday, September 10, 2022

### Data Sharing and Access

#### Prevalence and Characteristics of Data Sharing Policies Across the Health Research Life Cycle: Funders, Ethics Committees, Trial Registries, Journals, and Data Repositories

Aidan C. Tan,<sup>1</sup> Sol Libesman,<sup>1</sup> Weber Liu,<sup>1</sup> Zijing Yang,<sup>1</sup> Rani R. Chand,<sup>1</sup> Kylie E. Hunter,<sup>1</sup> Angela Webster,<sup>1</sup> Anna Lene Seidler<sup>1</sup>

**Objective** Most principal investigators support the concept of data sharing in principle, but few commit to sharing data in practice.<sup>1</sup> One way shown to reduce this gap is if major stakeholders across the research life cycle implement policies to recommend or require data sharing. The objective of this study was to determine the prevalence and characteristics of data sharing policies.

**Design** This was a cross-sectional study of data sharing policies of health research funders, research ethics committees, clinical trial registries, peer-reviewed scientific journals, and research data repositories. It included the 55 largest private and 55 largest public and philanthropic health research funders by annual health research expenditure, all national ethics committees, all clinical trial registries, the 5 highest-impact peer-reviewed scientific journals by journal impact factor for each of the 59 fields of clinical medicine, and all research data repositories in clinical medicine. Investigators reviewed all official websites, online reports, and gray literature information sources of stakeholders for the presence of a data sharing policy. If present, investigators assessed its magnitude of support for data sharing. If it recommended or required data sharing, investigators assessed its characteristics. All data were abstracted in duplicate by 2 independent investigators who compared the relevant information against structured criteria on a pre piloted data extraction form and resolved disagreements by discussion and a third investigator.

**Results** Overall, 110 health research funders, 124 national ethics committees, 18 clinical trial registries, 273 peer-reviewed scientific journals, and 410 research data repositories were included. More than half of health research funders either recommended (15 [15%]) or required (45 [41%]) data sharing. These policies typically applied to all

data from only interventional studies, with justified exceptions, and specified data to be shared before a predetermined period with independent committee-approved investigators for research proposal–approved purposes via third-party websites. Only 4 national ethics committees (3%) recommended data sharing. These policies typically applied to all studies, with justified exceptions, and specified data to be shared via third-party websites. Only 1 clinical trial registry (6%) required data sharing. This policy applied to only interventional studies, with justified exceptions, and specified data to be shared via third-party websites. Almost two-thirds of peer-reviewed scientific journals either recommended (120 [44%]) or required (52 [19%]) data sharing. These policies typically applied to only some data from all studies, with unjustified exceptions, and specified data to be shared with anyone for any purpose via third-party websites. Few research data repositories recommended (26 [6%]) or required (24 [6%]) data sharing. These policies typically applied to all data from all studies, with unjustified exceptions, and specified data to be shared with anyone for any purpose and via third-party websites.

**Conclusions** Data sharing imperatives were not met by most stakeholders.

## Reference

1. Tan AC, Askie LM, Hunter KE, Barba A, Simes RJ, Seidler AL. Data sharing—trialists' plans at registration, attitudes, barriers and facilitators: a cohort study and cross-sectional survey. *Res Synth Methods*. 2021;12(5):641-657. doi:10.1002/jrsm.1500

<sup>1</sup>NHMRC Clinical Trials Centre, The University of Sydney, Sydney, Australia, aidan.tan@sydney.edu.au

## Frequency of Data and Code Sharing in Medical Research: An Individual Participant Data Meta-analysis of Metaresearch Studies

Daniel G. Hamilton,<sup>1,2</sup> Kyungwan Hong,<sup>3</sup> Hannah Fraser,<sup>1</sup> Anisa Rowhani-Farid,<sup>3</sup> Steve McDonald,<sup>4</sup> Fiona Fidler,<sup>1,5</sup> Matthew J. Page<sup>4</sup>

**Objective** Numerous metaresearch studies have investigated rates and predictors of data and code sharing in medicine. However, these studies have often been narrow in scope, focusing on some important aspects and predictors of sharing but not others. A systematic review and individual participant data (IPD) meta-analysis of this corpus of research is being conducted to provide an expansive picture of how availability rates have changed over time in medicine and what factors are associated with sharing.

**Design** Ovid Embase, Ovid MEDLINE, MetaArXiv, medRxiv, and bioRxiv were searched up to July 1, 2021, for metaresearch studies that investigated data sharing, code sharing, or both among a sample of scientific articles presenting original research from the medical and health sciences (ie, primary articles). Two authors independently screened records and assessed risk of bias in the included studies. Key outcomes of interest included the prevalence of

affirmative sharing declarations (declared availability) and availability as confirmed by the metaresearch authors (actual availability). The association between data and code availability and several factors (eg, year published, journal policy) were also examined. IPD were collected or requested from authors of eligible studies. A 2-stage approach to IPD meta-analysis was performed, with outcomes pooled using the Hartung-Knapp-Sidik-Jonkman method for random-effects meta-analysis.<sup>1</sup> The review methods were preregistered on the Open Science Framework<sup>2</sup> and are described in a detailed review protocol.<sup>3</sup>

**Results** A total of 4970 potential studies were identified, of which 101 were eligible for the review, 28 of which did not publicly share any IPD. Eligible studies examined a median (IQR) of 203 (125-398) primary articles published between 1987 and 2020 across 32 unique medical disciplines. To date, data from 36 studies (including 7750 primary articles) have been processed. Only 1 study was classified as low risk of bias. Meta-analysis revealed declared and actual data availability rates of 9% (95% CI, 6%-14%; 23 studies) and 3% (95% CI, 1%-6%; 26 studies), respectively, since 2015, with no significant differences between rates when compared with the preceding 5-year period. The same finding was also noted for code sharing (all <1%). Early results also indicate that only 35% (95% CI, 18%-55%; 5 studies) and 16% (95% CI, 10%-22%; 2 studies) of authors complied with mandatory data and code sharing policies, respectively. Comparatively, 13% (95% CI, 0-37%; 6 studies) and 8% (95% CI, 0-50%; 4 studies) of authors submitting to journals with policies encouraging sharing or no policy made data available, respectively.

**Conclusions** Preliminary analysis suggests that data and code sharing in medicine remains uncommon and occurs at a rate much lower than expected if journal policies were followed. We recommend future research to explore why sharing rates and compliance with mandatory policies are low as well as strategies for how this might be improved.

## References

1. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. 2014;14(1):25. doi:10.1186/1471-2288-14-25
2. Hamilton DG, Fraser H, Fidler F, Rowhani-Farid A, Hong K, Page MJ. Rates and predictors of data and code sharing in the medical and health sciences: a systematic review and individual participant data meta-analysis. Open Science Framework. May 28, 2021. doi:10.17605/OSF.IO/7SX8U
3. Hamilton DG, Fraser H, Fidler F, et al. Rates and predictors of data and code sharing in the medical and health sciences: protocol for a systematic review and individual participant data meta-analysis. *F1000Research*. 2021;10:491. doi:10.12688/f1000research.53874.2

<sup>1</sup>MetaMelb Research Group, School of BioSciences, The University of Melbourne, Parkville, Australia, hamilton.d@unimelb.edu.au; <sup>2</sup>Melbourne Medical School, Faculty of Medicine, Dentistry &

Health Sciences, The University of Melbourne, Parkville, Australia; <sup>3</sup>Department of Pharmaceutical Health Services Research, University of Maryland, Baltimore, MD, USA; <sup>4</sup>School of Public Health & Preventive Medicine, Monash University, Melbourne, Australia; <sup>5</sup>School of Historical and Philosophical Studies, The University of Melbourne, Parkville, Australia

**Conflict of Interest Disclosures** Daniel G. Hamilton is a board member of the Association of Interdisciplinary Meta-research and Open Science (AIMOS) and a PhD candidate supported by an Australian Commonwealth Government Research Training Program Scholarship. The Laura and John Arnold Foundation funds the Restoring Invisible and Abandoned Trials (RIAT) Support Center, which supports the salary of Kyungwan Hong and Anisa Rowhani-Farid. Kyungwan Hong was supported in 2020 by the US Food and Drug Administration (FDA) of the US Department of Health and Human Services (HHS) as part of a financial assistance award UoIFD005946, funded by FDA/HHS. The project contents are those of Kyungwan Hong and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the US government.

### Assessment of Concordance Between Yale Open Data Access (YODA) Project Data Requests and Corresponding Publications

Enrique Vazquez,<sup>1</sup> Joseph S. Ross,<sup>2,3,4</sup> Cary P. Gross,<sup>2,5,6</sup> Karla Childers,<sup>7</sup> Stephen Bamford,<sup>8</sup> Joanne Waldstreicher,<sup>7</sup> Harlan M. Krumholz,<sup>3,4,9</sup> Joshua D. Wallach<sup>10</sup>

**Objective** The Yale Open Data Access (YODA) Project enables researchers to access shared participant-level clinical research data for independent secondary and replication studies.<sup>1</sup> The project, because it requires an application, provides an opportunity to determine how published analyses compare with the initial aims and the degree to which any deviance is noted in the publications. Accordingly, the objective of this study was to evaluate the concordance among the included trials, the study objectives, and the statistical methods specified in researchers' requests to the YODA Project for Johnson & Johnson clinical trial data, the primary YODA data sharing partner, and their corresponding publications.

**Design** In this cross-sectional study, all approved YODA requests for Johnson & Johnson pharmaceutical or medical device data that had 1 corresponding English-language publication or more were identified (from the first request in 2018 to October 29, 2021). From each request-publication pair, the primary objectives were classified as fully, partially, or not at all concordant. Primary and secondary end points were classified as fully concordant, partially concordant ( $\geq 1$  additional primary or secondary end point in the request or publication), or discordant ( $\geq 1$  secondary end point dropped or converted to a primary end point, primary end point converted to a secondary end point, or secondary and primary end points swapped). Given that slight methodological changes may have been necessary once researchers had access to the shared data, statistical methods were classified as concordant if the pairs described the same broad methodological approaches.

**Table 17. Concordance Between Yale Open Data Access Project Data Requests and Corresponding Publications**

Characteristic	Request-publication pairs, No. (%) (N = 48)
<b>Study objective(s)</b>	
Fully concordant	33 (68.8)
Partially concordant	13 (27.1)
Discordant	2 (4.2)
<b>No. of studies requested and analyzed</b>	
Fully concordant	28 (58.3)
Discordant	20 (41.7)
Greater No. of studies listed in the data request	17 (35.4)
Greater No. of studies listed in the publication	0
Unclear No. of studies in the publication	3 (6.3)
<b>Primary end point(s)</b>	
Fully concordant	31 (64.6)
Partially concordant	7 (14.6)
Additional primary end point(s) in data request	4 (8.3)
Additional primary end point(s) in publication	3 (6.25)
Discordant <sup>a</sup>	10 (20.8)
<b>Secondary end point(s)</b>	
Fully concordant	25 (52.1)
Partially concordant	11 (22.9)
Additional secondary end point(s) in data request	3 (6.3)
Additional secondary end point(s) in publication	8 (16.7)
Discordant <sup>a</sup>	12 (25.0)
<b>Statistical methods</b>	
Concordant	39 (81.3)
Discordant	9 (18.8)

<sup>a</sup>At least 1 primary end point dropped or converted to a secondary end point, secondary end point dropped or converted to a primary end point, or primary and secondary end points swapped.

**Results** Forty-eight requests on the YODA Project website with 1 publication or more in a peer-reviewed journal were identified. Of the 48 request-publication pairs, 33 (68.8%) had a fully concordant overarching study objective, and 13 (27.1%) had a partially concordant overarching study objective (Table 17). There were 28 pairs (58.3%) for which all of the requested trials were included in the analyses described in the publications; 17 pairs (35.4%) had articles that included fewer trials than the number of trials specified in the request. There were 31 pairs (64.6%) with fully concordant primary end points and 25 pairs (52.1%) with fully concordant secondary end points. Only 1 pair had fully concordant primary and secondary end points. Most pairs (39 [81.3%]) had concordant statistical methods; there were no pairs that were fully concordant across all proposal details.

**Conclusions:** None of the YODA Project requests were fully concordant with their corresponding publications describing the completed research, most often because fewer trials were used than requested. These findings suggest that investigators using data from data sharing platforms should explain

deviations from the data requests in their publications and that research reviewers should compare and evaluate the consistency between the prespecified requests and publications.

## Reference

1. Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Sci Data*. 2018;5(1):180268. doi:10.1038/sdata.2018.268

<sup>1</sup>Yale University, New Haven, CT, USA; <sup>2</sup>Section of General Internal Medicine, Yale School of Medicine, New Haven, CT, USA; <sup>3</sup>Yale–New Haven Hospital Center for Outcomes Research and Evaluation, New Haven, CT, USA; <sup>4</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA; <sup>5</sup>Cancer Outcomes, Public Policy, and Effectiveness Research (COPPER) Center, Yale University, New Haven, CT, USA; <sup>6</sup>Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA; <sup>7</sup>Johnson & Johnson, New Brunswick, NJ, USA; <sup>8</sup>Johnson & Johnson, High Wycombe, UK; <sup>9</sup>Section of Cardiovascular Medicine, Yale School of Medicine, New Haven, CT, USA; <sup>10</sup>Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA, joshua.wallach@yale.edu

**Conflict of Interest Disclosures** Joseph S. Ross and Harlan M. Krumholz report being cofounders of the Yale Open Data Access (YODA) Project, and Cary P. Gross and Joshua D. Wallach report being YODA project affiliates. Joseph S. Ross is a former associate editor of *JAMA Internal Medicine* and a current research editor at *The BMJ* and receives research support through Yale University from Johnson & Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology, from the US Food and Drug Administration (FDA) for the Yale–Mayo Clinic Center for Excellence in Regulatory Science and Innovation program (grant U01FD005938), from the Agency for Healthcare Research and Quality (grant R01HS022882), from the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) (grants R01HS025164 and R01HL144644), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International; in addition, he is an expert witness at the request of the relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. Cary P. Gross has received research funding through Yale University from the National Comprehensive Cancer Network Foundation (funded by AstraZeneca) and Johnson & Johnson to help devise and implement new approaches to sharing clinical trial data and from Genentech. Karla Childers, Stephen Bamford, and Joanne Waldstreicher are employees and stockholders of Johnson & Johnson. Harlan M. Krumholz reports that he has contracts through Yale New Haven Hospital with the Centers for Medicare & Medicaid Services to support quality measurement programs and through Yale University with UnitedHealth Group to engage in collaborative research. He was a recipient of a research grant through Yale University from Medtronic for data sharing, from the FDA to develop methods for postmarket surveillance of medical devices, from Johnson & Johnson to support data sharing, and from the Shenzhen Center for Health Information for work to advance intelligent disease prevention and health promotion; he is an advisor to the National Center for Cardiovascular Diseases in Beijing, China; was an expert witness for the Arnold & Porter Law Firm for work related to the Sanofi clopidogrel litigation; and is an expert witness for the Martin/Baughman Law Firm for work related to the Cook Celect inferior vena cava (IVC) filter litigation and related to C. R. Bard Recovery IVC filter litigation and for the Siegfried and Jensen Law Firm for work related to Vioxx litigation; he chairs a cardiac scientific advisory board for UnitedHealth; was a

member of the IBM Watson Health Life Sciences Board; is a member of the advisory board for Element Science, the health care advisory board for Facebook, and the physician advisory board for Aetna; he is the cofounder of HugoHealth, a personal health information platform, and cofounder of Refactor Health, an enterprise health care artificial intelligence–augmented data management company; he is a venture partner at F-Prime. Joshua D. Wallach is supported by the FDA and the National Institute on Alcohol Abuse and Alcoholism of the NIH under award 1K01AA028258.

## Sharing of Individual Participant-Level Data by Trialists of Randomized Clinical Trials of Pharmacologic Treatments for COVID-19

Laura Esmail,<sup>1,2,3</sup> Philipp Kapp,<sup>4</sup> Rouba Assi,<sup>1,2,3</sup> Julie Wood,<sup>5</sup> Gabriela Regan,<sup>5</sup> Philippe Ravaud,<sup>1,2,3</sup> Isabelle Boutron<sup>1,2,3</sup>

**Objective** The COVID-19 pandemic may have signaled a positive shift in attitudes toward sharing individual-level patient data (IPD).<sup>1</sup> This project aimed to obtain IPD from trialists of randomized clinical trials (RCTs) of pharmacological treatments for COVID-19 for the purposes of conducting an IPD meta-analysis under the COVID-NMA initiative.<sup>2</sup>

**Design** This single cohort study evaluated the effectiveness of accessing IPD from trialists of COVID-19 RCTs through email requests and online data repositories. Participants were the corresponding authors of RCTs of pharmacological trials for the treatment of COVID-19 who published their findings in a preprint or peer-reviewed journal between March 2020 and May 2021. Corresponding authors were emailed once, with at least 2 reminders over 9 months (November 2020 to September 2021) regardless of their data sharing statement. Online data repositories (including Vivli, Yale University Open Data Access Project, and ClinicalStudyDataRequest.com) were also searched. The primary outcome was the proportion of studies for which IPD were accessed; IPD was defined as a data set obtained directly from trialists or those that were made accessible online. The project team collaborated with Vivli, a global data sharing platform.

**Results** Fifty-six of 229 COVID-19 RCT trialists (24%) shared their IPD by the end of December 2021. Of these, 18 (32%) declared their positive willingness to share in the registry and 42 (75%) in their preprint or publication. Of those trials that did not share their data, 53 (31%) declared their positive willingness to share in the registry and 100 (58%) in their preprint or publication. Stratified by funding, trials that shared vs did not share IPD received funding from sources that were public or nonprofit (shared: 13 [23%]; not shared: 79 [46%]), private (shared: 10 [18%]; not shared: 36 [21%]), mixed public/nonprofit and private (shared: 20 [36%]; not shared: 37 [21%]), no funding (shared: 11 [20%]; not shared: 10 [6%]), and not reported or unclear (shared: 2 [4%]; not shared: 11 [6%]). Trials that shared vs did not share IPD were based in high-income countries (shared: 19 [34%]; not shared: 45 [26%]), low- to middle-income countries (shared: 31 [55%]; not shared: 117 [68%]), and both high-income and low- to middle-income countries (shared: 6 [11%]; not shared: 11 [6%]).

**Conclusions** Despite the positive shift in attitudes toward sharing IPD, less than one-quarter of trialists shared their IPD. Furthermore, data sharing statements often did not line up with the ability to obtain the data. This study emphasizes the need to mandate and/or reward timely data sharing while addressing remaining administrative, resource infrastructure, and cultural obstacles.

## References

1. Larson K, Sim I, von Isenburg M, et al. COVID-19 interventional trials: analysis of data sharing intentions during a time of pandemic. *Contemp Clin Trials*. 2022;115:106709. doi:10.1016/j.cct.2022.106709
2. Boutron I, Chaimani A, Meerpohl JJ, et al; COVID-NMA Consortium. The COVID-NMA project: building an evidence ecosystem for the COVID-19 pandemic. *Ann Intern Med*. 2022;173(12):1015-1017. doi:10.7326/M20-5261

<sup>1</sup>Université de Paris, INSERM, INRAE, CNAM, Centre de Recherche Épidémiologie et Statistiques, Paris, France, laura.esmail@healthevidence.eu; <sup>2</sup>Centre d'Épidémiologie Clinique, Assistance Publique Hôpitaux de Paris, Hôpital Hôtel-Dieu, Paris, France; <sup>3</sup>Cochrane France, Paris, France; <sup>4</sup>Institute for Evidence in Medicine, Medical Center–University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany; <sup>5</sup>Vivli, Cambridge, MA, USA

**Conflict of Interest Disclosures** Isabelle Boutron is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. No other disclosures were reported.

**Funding/Support** This study, as part of the COVID-NMA initiative, received funding from Université de Paris, Assistance Publique Hôpitaux de Paris, INSERM, Cochrane France (Ministry of Health), the French Ministry of Higher Education and Research, Agence Nationale de la Recherche, and the World Health Organization.

**Additional Information** Isabelle Boutron is a co-corresponding author.

## Preprints

### medRxiv Preprint Submissions, Posts, and Key Metrics, 2019-2021

Joseph S. Ross,<sup>1</sup> Richard Sever,<sup>2</sup> Theodora Bloom,<sup>3</sup> Samantha Hindle,<sup>2</sup> Dinar Yunusov,<sup>2</sup> Theodore Roeder,<sup>2</sup> John R. Inglis,<sup>2</sup> Harlan M. Krumholz<sup>4</sup>

**Objective** Preprint servers offer a means to disseminate research reports before or concurrent with peer-review.<sup>1</sup> medRxiv, an independent, not-for-profit preprint server for clinical and health science research introduced in June 2019, grew substantially in the context of the COVID-19 pandemic.<sup>2</sup> Submissions, preprints posted, and user downloads for medRxiv since launch were characterized.

**Design** This analysis used data from the medRxiv website, internal data, and Altmetric.com from June 11, 2019 (launch), through December 31, 2021. Submissions, postings, abstract views, downloads, comments, and withdrawals were assessed.

The posting rate was calculated as the percentage of submissions that were posted after passing screening criteria, including that the submission represents scientific research (not a narrative review, commentary, or case report). In addition, all posted preprints with Altmetric scores greater than 1000 were identified. Published journal articles corresponding to posted preprints were identified through routine, automated searches of PubMed and CrossRef.

**Results** As of December 31, 2021, there were 33,342 submissions to medRxiv, 27,674 (83.0%) of which were subsequently posted after screening: 913 in 2019, 14,070 in 2020, and 12,691 in 2021. Among these, 6165 (22.3%) had been revised at least once and 4227 (15.3%) were simultaneously submitted to journals as part of the M2J program. Overall, 16,465 preprints (59.5%) described COVID-19 research. In total, 47 posted preprints (0.17%) were subsequently withdrawn, 30 of which were COVID-19-related. Preprints have thus far been posted by 156,290 unique authors from 151 countries, most commonly from the United States, the United Kingdom, and China. As of December 31, 2021, there were 51,943,342 downloads and 132,900,392 abstract views: 107,772 and 241,528 in 2019; 27,963,915 and 61,613,928 in 2020; and 23,871,655 and 71,044,936 in 2021. The median (IQR) number of downloads per preprint was 451 (261-904) and abstract views per preprint was 1659 (1055-2955). There have been 8394 total user comments on preprints and 2199 preprints (7.9%) have at least 1 user comment. There were 312 preprints (1.1%) with an Altmetric score greater than 1000, all but 1 of which was COVID-19-related; the median (IQR) Altmetric score was 2 (0-11). Thus far, 10,041 preprints (36.3%) were subsequently published in 2316 peer-reviewed journals, with a median interval between preprint posting and journal publication of 140 days, including 566 (62.0%), 6615 (47.0%), and 2860 (22.5%) that were published after being posted in 2019, 2020, and 2021, respectively.

**Conclusions** medRxiv grew rapidly since its launch, particularly for COVID-19-related research. The preprint server is an active repository for clinical and health science research; future research should also account for peer communication through social media.

## References

1. Krumholz HM, Ross JS, Otto CM. Will research preprints improve healthcare for patients? *BMJ*. 2018;362:k3628. doi:10.1136/bmj.k3628
2. Krumholz HM, Bloom T, Sever R, et al. Submissions and downloads of preprints in the first year of medRxiv. *JAMA*. 2020;324:1903-1905. doi:10.1001/jama.2020.17529

<sup>1</sup>Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA, joseph.ross@yale.edu; <sup>2</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; <sup>3</sup>The BMJ, London, UK; <sup>4</sup>Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

**Conflict of Interest Disclosures** All authors disclose being cofounders of medRxiv. Joseph Ross is a former associate editor of *JAMA Internal Medicine*, a current research editor at *The BMJ*, and receives research support through Yale University from Johnson & Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology (NEST), from the Food and Drug Administration for the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI) program (U01FD005938), from the Agency for Healthcare Research and Quality (R01HS022882), from the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International; in addition, Joseph Ross is an expert witness at the request of relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. Richard Sever reports being the assistant director of Cold Spring Harbor Laboratory Press and a director of Life Science Alliance LLC. Theodora Bloom is employed full time by *The BMJ*; reports chairing the scientific advisory board of EMBL-EBI Literature Services (<http://www.ebi.ac.uk/services/literature>), being on the Board of Managers of AIP Publishing (<https://publishing.aip.org/>), and being European coordinator for the Peer Review Congress (<https://peerreviewcongress.org/organizers-and-board.html>). Samantha Hindle is cofounder of PREreview, an initiative to support and train early-career researchers in peer review using preprints. John R. Inglis reports being the executive director and publisher of Cold Spring Harbor Laboratory Press, a director of Life Science Alliance LLC, and a member of the advisory board of MIT Press. Harlan M. Krumholz reports that he is a cofounder of Refactor Health and HugoHealth, is associated with contracts, through Yale New Haven Hospital from the Centers for Medicare & Medicaid Services and through Yale University from Johnson & Johnson, and has received expenses and/or personal fees from UnitedHealth, Element Science, Aetna, Reality Labs, Tesseract/4Catalyst, the Siegfried and Jensen Law Firm, Arnold and Porter Law Firm, Martin/Baughman Law Firm, and F-Prime. Theodora Bloom is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

**Funding/Support** This work was supported in part by funds from Chan Zuckerberg Initiative through the Cold Spring Harbor Laboratory.

**Role of the Funder/Sponsor** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Assessment of Concordance Between Reports of Clinical Studies Posted as medRxiv Preprints and Corresponding Publications in Peer Reviewed Journals

Guneet Janda,<sup>1</sup> Vishal Khetpal,<sup>2</sup> Xiaoting Shi,<sup>3</sup> Joseph S. Ross,<sup>4,5,6</sup> Joshua D. Wallach<sup>7</sup>

**Objective** The study characteristics, results, and interpretations described in preprints of clinical studies that are subsequently published in high-impact journals are broadly concordant.<sup>1</sup> Given that studies published in high-impact journals may represent the highest-quality studies, it is necessary to evaluate concordance for a larger sample of clinical studies posted as preprints and subsequently published in any journal, regardless of impact factor (IF).

**Design** In this cross-sectional analysis, preprints posted on medRxiv in September 2020 were identified. Four evaluators determined how many preprints were subsequently published in peer-reviewed journals as of March 2022, calculating the time from preprint posting to publication. For preprints with multiple versions, the most recent version prior to journal acceptance was selected. Preprints updated after journal acceptance were excluded. For preprint-journal article pairs describing clinical trials, observational studies, and meta-analyses that measured health-related outcomes, sample size, primary end points, corresponding results, and overarching conclusions were abstracted. Results from primary end points were considered concordant only if they contained the same information or had numerical equivalence (eg, identical effect size estimates and 95% CIs or *P* values from inferential analyses). Rates of concordance were compared between preprints and corresponding journal articles overall and by focus on COVID-19 and journal IF.

**Results** Among 1399 preprints first posted on medRxiv in September 2020, there were 623 modeling studies (44.5%), 280 observational studies (20.0%), 90 systematic reviews or meta-analyses (6.4%), 42 clinical trials (3.0%), and 364 articles with other study designs (26.0%). As of March 2022, there were 680 preprints (48.6%) published a median (IQR) of 5 (3-7) months after preprint posting. Among 331 preprint-publication pairs describing clinical trials, observational studies, or meta-analyses, 182 pairs (55.0%) were related to COVID-19. Of 325 pairs reporting sample sizes in both sources, 290 pairs (89.2%) were concordant. Of 35 pairs with discordant sample sizes, 20 pairs (57.1%) had larger samples in the journal publication. There were 328 pairs (99.1%) with concordant and 3 pairs (0.8%) with discordant primary end points. Among 329 pairs in which results for primary end points could be compared, 290 pairs (88.1%) were concordant. Two-thirds of 39 discordant pairs (26 pairs [66.7%]) had effect size estimates in the same direction and were statistically consistent. Overall, 323 pairs (97.6%) had concordant study interpretations, including 32 of 39 pairs with discordant primary end point results (82.1%). Pairs with corresponding publications in journals with an IF of 10 or higher had lower concordance rates for sample size (17 of 24 pairs [70.8%] vs 235 of 258 pairs [91.1%]; *P* = .01) and results (16 of 23 pairs [70.0%] vs 235 of 261 pairs [90.0%]; *P* = .004).

**Conclusions** Most clinical studies posted as preprints on medRxiv and subsequently published in peer-reviewed journals had concordant study characteristics, results, and interpretations, similar to what has been previously observed among preprints published in the highest-impact journals.<sup>1</sup>

## Reference

- Shi X, Ross JS, Amancharla N, Niforatos JD, Krumholz HM, Wallach JD. Assessment of concordance and discordance among clinical studies posted as preprints and subsequently published in high-impact journals. *JAMA Netw Open*. 2021;4(3):e212110. doi:10.1001/jamanetworkopen.2021.2110

<sup>1</sup>Yale School of Medicine, New Haven, CT, USA; <sup>2</sup>Department of Medicine, Warren Alpert Medical School of Brown University, Providence, RI, USA; <sup>3</sup>Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA; <sup>4</sup>Section of General Internal Medicine, Yale School of Medicine, New Haven, CT, USA; <sup>5</sup>Yale-New Haven Hospital Center for Outcomes Research and Evaluation, New Haven, CT, USA; <sup>6</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA; <sup>7</sup>Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA, joshua.wallach@yale.edu

**Conflict of Interest Disclosures** Guneet Janda receives research funding from the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI) Scholars Award Program. Vishal Khetpal serves in an advisory role for Necessary Ventures. Xiaoting Shi is supported by the China Scholarship Council. Joseph S. Ross is a former associate editor of *JAMA Internal Medicine* and a current research editor at *The BMJ*; receives research support through Yale University from Johnson & Johnson to develop methods of clinical trial data sharing, the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology, the US Food and Drug Administration (FDA) for CERSI, the Agency for Healthcare Research and Quality, the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH), and the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International; and is an expert witness at the request of relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and the Anti-Kickback Statute against Biogen. Joseph S. Ross reported cofounding medRxiv, the preprint server for health sciences research, and Joshua D. Wallach reported being a medRxiv affiliate; these individuals did not receive compensation for their roles. Joshua D. Wallach is supported by the National Institute on Alcohol Abuse and Alcoholism of the NIH.

### Comparison of Reports of Epidemiology Studies Posted as bioRxiv Preprints and Published in Peer Reviewed Journals

Mario Malički,<sup>1</sup> Ana Jerončić,<sup>2</sup> Gerben ter Riet,<sup>3,4</sup> Lex Bouter,<sup>5,6</sup> John P. A. Ioannidis,<sup>1,7,8,9,10</sup> IJsbrand Jan Aalbersberg,<sup>11</sup> Steven N. Goodman<sup>1,7,8</sup>

**Objective** Previous studies showed high levels of similarity between preprints and their subsequent peer-reviewed journal publications.<sup>1,2</sup> The goal of this study was to analyze the extent of similarity for preprint-publication pairs in the field of epidemiology.

**Design** This cohort study documented differences between bioRxiv epidemiology preprints with only 1 preprint version and their subsequent journal version. Preprints were classified as “epidemiology” by their submitting authors. From inception of the preprint server through December 31, 2019, there were 622 such preprints. Sample size calculation for precision using a 95% confidence level and an 8% margin of error yielded a requirement of 121 preprints, which were then randomly sampled from the 622. Changes between preprint-publication pairs were highlighted using the Microsoft Word function *compare two versions of a document*. Any changes that occurred were noted and classified.

**Results** The 121 bioRxiv epidemiology preprints were later published in 73 different journals (median [IQR] impact

factor, 4 [2.9–6.9]) with a median (IQR) time from preprint to publication of 204 (131–243) days. Of the 121 pairs, 31 (26%) had differences in their titles, 8 (7%) in the number or order of authors, 31 (26%) in the number of tables, 28 (23%) in the number of figures, 102 (84%) in the number of references, 54 (44%) in acknowledgment descriptions, 74 (61%) in conflict of interest declarations, and 49 (40%) in data sharing statements. Regarding main content, 109 (90%) had changes in the abstract, 7 of which (6%) reported different *P* values; 106 (88%) had changes in the introduction section, 37 of which (31%) altered descriptions of their objectives; 120 (99%) had changes in their methods section, 9 of which (7%) had changes in their sample size; 115 (95%) had changes in their results section, with 82 (68%) adding or removing (parts of) results; and 116 (96%) had changes in the their discussion sections, with 65 (54%) adding limitations in their journal versions and 12 (10%) exhibiting substantive changes to main results in the first sentence of their discussion (**Table 18**).

**Conclusions** This study shows that almost all aspects of epidemiological preprints were slightly changed in their journal publication versions, with 10% of preprints changing their main findings. Further research is needed to determine who requested those changes and why, whether changes were associated with the quality of the study or the expertise of those requesting them, and whether changes led to increases in validity, transparency, or readability.

### References

- Klein M, Broadwell P, Farb SE, Grappone T. Comparing published scientific journal articles to their pre-print versions. *Int J Digital Librar.* 2019;20:335–350. doi:10.1007/s00799-018-0234-1
- Shi X, Ross JS, Amancharla N, Niforatos JD, Krumholz HM, Wallach JD. Assessment of concordance and discordance among clinical studies posted as preprints and subsequently published in high-impact journals. *JAMA Netw Open.* 2021;4(3):e212110. doi:10.1001/jamanetworkopen.2021.2110

**Table 18. Changes Between Epidemiology Studies Posted as bioRxiv Preprints and Published in Peer-Reviewed Journals**

Section with changes	No. (%) (N = 121)
Title	31 (26)
Authorship	8 (7)
Abstract: <i>P</i> values	7 (6)
Introduction: objectives	9 (7)
Methods: sample size	31 (26)
Results: addition or removal of parts of results	82 (68)
Discussion: main results finding	12 (10)
Statements	
Acknowledgments	54 (44)
Conflicts of interest	74 (61)
Data sharing	49 (40)
References	102 (84)

<sup>1</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA, mario.malicki@mefst.hr; <sup>2</sup>Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia; <sup>3</sup>Urban Vitality Centre of Expertise, Amsterdam University of Applied Sciences, Amsterdam, the Netherlands; <sup>4</sup>Amsterdam University Medical Centers, Department of Cardiology, Amsterdam, the Netherlands; <sup>5</sup>Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, the Netherlands; <sup>6</sup>Amsterdam University Medical Centers, Department of Epidemiology and Data Science, Amsterdam, the Netherlands; <sup>7</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; <sup>8</sup>Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA; <sup>9</sup>Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA; <sup>10</sup>Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA; <sup>11</sup>Elsevier, Amsterdam, the Netherlands

**Conflict of Interest Disclosures** IJsbrand Jan Aalbersberg is senior vice president of research integrity at Elsevier. Mario Malički is a co-editor in chief of *Research Integrity and Peer Review*. Lex Bouter, John P. A. Ioannidis, and Steven N. Goodman are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract.

**Funding/Support** Elsevier funding was awarded to Stanford University for a Meta-Research Innovation Center at Stanford postdoctoral position that supported Mario Malički's work on the project.

**Role of the Funder/Sponsor** IJsbrand Jan Aalbersberg is an employee of Elsevier and had a role in the design and conduct of the study; management and interpretation of the data; review and approval of the abstract; and decision to submit the abstract for presentation.

### Content Analysis of Comments on bioRxiv and medRxiv Preprints

Clarissa França Dias Carneiro,<sup>1,2</sup> Danielle Rayêe,<sup>3</sup> Flávia Boos,<sup>4</sup> Gabriel Costa,<sup>1</sup> Kleber Neves,<sup>1</sup> Mariana Boechat de Abreu,<sup>5</sup> Pedro Batista Tan,<sup>1</sup> Roberta Andrejew,<sup>6</sup> Tiago Lubiana,<sup>7,8</sup> Mario Malički,<sup>9</sup> Olavo B. Amaral<sup>1</sup>

**Objective** This study aimed to describe and explore the content of comments received by preprints posted in 2020 to bioRxiv and medRxiv. It extended a previous study<sup>1</sup> that assessed comments on bioRxiv until 2019, considering the increased attention received by preprints after the beginning of the COVID-19 pandemic.<sup>2</sup>

**Design** This was an observational qualitative study. The list of preprints posted to bioRxiv and medRxiv and the number of comments received by them were obtained using each platform's application program interface. After excluding preprints with no comments and those with more than 20 comments (for feasibility and avoiding overrepresentation of individual preprints), eligible preprints were randomly sampled and their comments were assigned to 3 reviewers. Content analysis of comments was based on a structured form, with predefined categories based on previous research on preprint comments<sup>1</sup> and on a scoping review of peer review.<sup>3</sup> If 2 of 3 reviewers agreed, the classification was considered final; in case all 3 disagreed, a fourth reviewer made the final call. Neither manuscripts nor the comment

**Table 19. Summary of Main Content Identified in Comments Posted in Biomedical Preprints**

Category	No. of comments identified (No. of applicable comments)	% of Comments (95% CI)
Comments responding to previous ones	439 (1921)	22.9 (21.0-24.7)
Comments from the authors	165 (1482)	11.1 (9.5-12.7)
Comments not about the preprint content	192 (1317)	14.6 (12.7-16.5)
Comments including any compliment or positive appraisal	428 (1125)	38.0 (35.2-40.9)
Comments including any criticisms, corrections, or suggestions	694 (1125)	61.7 (58.8-64.5)
Comments including any questions	393 (1125)	34.9 (32.1-37.7)

appropriateness or validity in regard to the manuscript text were assessed.

**Results** A total of 1037 preprints were assessed, with a total of 1921 comments (mean [SD], 1.8 [1.9] comments per preprint); 439 comments (22.8%) were replies to previous comments, leaving 1482 comments available for classification based on our prespecified categories (**Table 19**). A total of 14.6% (95% CI, 12.7%-16.5%) of reader comments were considered not to be about the content of the preprints (n = 192). Comments from preprint authors (n = 165) accounted for 11.1% (95% CI, 9.5%-12.7%) of the remaining comments, and mostly addressed publication status (n = 89); 38% (95% CI, 35.2%-40.9%) of reader (nonauthor) comments included compliments (n = 428), 61.7% (95% CI, 58.8%-64.5%) included criticisms, corrections, or suggestions (n = 694), and 34.9% (95% CI, 32.1%-37.7%) included questions (n = 393). Although compliments were mostly general (n = 279), criticisms, corrections, or suggestions largely addressed specific points, such as interpretation (n = 286), methodologic design (n = 267), materials/data collection methods (n = 238), and analyses (n = 228). Most questions were about information not provided in the preprint (n = 170) or about materials/data collection (n = 166).

**Conclusions** As shown before,<sup>1</sup> only a small percentage of preprints received comments on their respective preprint platforms, but this study found that the content of these comments showed similarities with the type of content typically expected from journal-elicited peer review as previously described.<sup>3</sup> These results may help generate hypotheses to inform future research on preprints and peer review.

### References

1. Malički M, Costello J, Alperin JP, Maggio LA. Analysis of single comments left for bioRxiv preprints till September 2019. *Biochem Med (Zagreb)*. Published online April 15, 2021. doi:10.11613/BM.2021.020201
2. Fraser N, Brierley L, Dey G, et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS*

*Biol.* Published online April 2, 2021. doi:10.1371/journal.pbio.3000959

3. Glonti K, Cauchi D, Cobo E, Boutron I, Moher D, Hren D. A scoping review on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. *BMC Med.* Published online June 20, 2019. doi:10.1186/s12916-019-1347-o

<sup>1</sup>Institute of Medical Biochemistry Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, clarissa.franca@bioqmed.ufrj.br, clarissa.franca-dias-carneiro@charite.de; <sup>2</sup>BIH QUEST Center for Transforming Biomedical Research, Berlin Institute of Health, Berlin, Germany; <sup>3</sup>Ophthalmology and Visual Sciences Department, Albert Einstein College of Medicine, New York, NY, USA; <sup>4</sup>Programa de pós-graduação em Psicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil; <sup>5</sup>Carlos Chagas Filho Institute of Biophysics, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; <sup>6</sup>Department of Biochemistry, Institute of Chemistry, Universidade de São Paulo, São Paulo, Brazil; <sup>7</sup>Ronin Institute, virtual organization; <sup>8</sup>Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, Brazil; <sup>9</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** Clarissa França Dias Carneiro received a scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Olavo B. Amaral received funding from FAPERJ (E-26/200.824/2021), CNPq (310813/2021-2), and the Serrapilheira Institute.

**Role of the Funder/Sponsor** The funders had no direct role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgment** The authors thank Richard Abdill for his contributions to data collection.

**Additional Information** Mario Malički and Olavo B. Amaral are co-corresponding authors.

## Media Attention, Twitter Engagement, and Citations of COVID-19 Clinical Trial Preprints and Their Corresponding Peer-Reviewed Publications

Emily Inwards,<sup>1</sup> Jennifer Klavens,<sup>1</sup> Amanda C. Adams,<sup>2</sup> Brian W. Roberts,<sup>1</sup> Timothy F. Platts-Mills,<sup>3</sup> Christopher W. Jones<sup>1</sup>

**Objective** To compare media attention, Twitter engagement, and citations among COVID-19–related clinical trial preprints and corresponding peer-reviewed publications.

**Design** Preprints of clinical trials were included in this cross-sectional study if they reported results related to the treatment or prevention of COVID-19 and were uploaded to an open access preprint server indexed by the National Institutes of Health iSearch COVID-19 portfolio in 2020. Two investigators, including a medical librarian, independently searched Medline, Google Scholar, and Embase to identify peer-reviewed publications corresponding to the included preprints. Altmetric data were used to quantify media mentions and Twitter interactions within 3 months after

publication for both preprints and peer-reviewed manuscripts. Citation counts were also recorded using Web of Science. Descriptive statistics were reported for the included trials, and linear regression was used to assess associations between study characteristics and media mentions, Twitter interactions, and citation counts.

**Results** Of 22,615 preprints screened for eligibility, 145 were included. MedRxiv was the source for most eligible preprints (n =100; 69%). Peer reviewed publications matching 118 of 145 preprints (81%) were found. Sixty-eight preprints (47%) received media citations within 3 months of publication (median [IQR] number of mentions per preprint, 0 [0-9]; maximum mentions, 568), and 118 (81%) had Twitter interactions (median [IQR] tweets per preprint, 28 [4-202]; maximum mentions, 18,177). One hundred preprints (69%) were cited in published literature (median [IQR] citations, 2 [0-12]). Among 118 preprints with matching peer-reviewed publications, Altmetric and citation data were unavailable for 5 (4%). Preprints received more media mentions than the corresponding peer-reviewed publications in 33 cases (28%), equal mentions in 38 cases (32%), and fewer media mentions in 42 cases (36%). Twitter mentions were greater for preprints than peer-reviewed publications in 54 cases (46%), equal in 2 (2%), and fewer for 62 (53%). In 26 cases (22%), preprints received more citations than peer-reviewed publications; citation counts were equal in 6 cases (5%) and in 85 cases (72%), the peer-reviewed version received more citations. Study size, government funding, failure to prospectively register, and positive study results were most often associated with increased media mentions, tweets, and citations (**Table 20**).

**Conclusions** Although peer-reviewed publications had more media, Twitter, and citation activity than corresponding preprints in most cases, it was not uncommon for preprints to receive more attention than peer-reviewed publications. Measures of trial reliability or quality were generally not associated with increases in these metrics.

<sup>1</sup>Department of Emergency Medicine, Cooper Medical School of Rowan University, Camden, NJ, USA, jones-christopher@cooperhealth.edu; <sup>2</sup>Medical Library, Cooper Medical School of Rowan University, Camden, NJ, USA; <sup>3</sup>Ophirex Inc, Corte Madera, CA, USA

**Conflict of Interest Disclosures** Christopher W. Jones reported receiving grants from AstraZeneca, Abbott, Vapotherm, and Ophirex outside the submitted work. Timothy F. Platts-Mills is an employee of Ophirex. No other disclosures were reported.

**Funding/Support** This work was supported by the US Department of Health and Human Services (HHS) Office of Research Integrity (grant ORIIIR200066-01-00).

**Role of the Funder/Sponsor** The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the abstract for presentation.

**Additional Information** The contents of this manuscript are those of the authors and do not represent the official views of, nor

**Table 20. Associations Between Trial Characteristics and Media Mentions, Tweets, and Citation Counts Among Preprints and Peer-Reviewed Publications<sup>a</sup>**

	Dependent variable					
	Preprint media mentions (n = 88)	Peer-reviewed publication media mentions (n = 73)	Preprint tweets (n = 88)	Peer-reviewed publication tweets (n = 73)	Preprint citations (n = 88)	Peer-reviewed citations (n = 73)
Government funding, regression coefficient (95% CI)	14 (-10 to 38)	56 (7 to 106) <sup>b</sup>	1137 (-75 to 2350)	909 (223 to 1594) <sup>b</sup>	45 (11 to 79) <sup>b</sup>	129 (-77 to 335)
Industry funding, regression coefficient (95% CI)	-1 (-30 to 28)	-0.2 (-58 to 58)	-108 (-1580 to 1363)	729 (-81 to 1539)	15 (-26 to 56)	37 (-206 to 280)
Registered prospectively, regression coefficient (95% CI)	-3 (-30 to 24)	-55 (-110 to -1) <sup>b</sup>	-1351 (-2702 to 0)	-918 (-1670 to -166) <sup>b</sup>	-44 (-82 to -7) <sup>b</sup>	-165 (-390 to 61)
Randomized allocation, regression coefficient (95% CI)	15 (-19 to 50)	73 (2 to 144)	1591 (-150 to 3333)	1042 (64 to 2020) <sup>b</sup>	37 (-11 to 85)	231 (-62 to 525)
Blinded participants, regression coefficient (95% CI)	-8 (-62 to 46)	4 (-110 to 119)	-368 (-3103 to 2367)	-50 (-1638 to 1538)	-8 (-84 to 68)	-106 (-583 to 370)
Blinded outcome assessors, regression coefficient (95% CI)	13 (-41 to 68)	46 (-66 to 159)	138 (-2594 to 2871)	304 (-1252 to 1859)	-5 (-81 to 71)	136 (-331 to 602)
No. of participants (in hundreds), regression coefficient (95% CI)	3 (3 to 4) <sup>b</sup>	2 (1 to 4) <sup>b</sup>	66 (32 to 100) <sup>b</sup>	38 (20 to 55) <sup>b</sup>	-1 (-1 to 0)	12 (7 to 17) <sup>b</sup>
Statistically significant result, regression coefficient (95% CI)	25 (1 to 48) <sup>b</sup>	75 (26 to 124) <sup>b</sup>	380 (-811 to 1570)	625 (-54 to 1304)	5 (-28 to 38)	262 (59 to 466) <sup>b</sup>
R <sup>2</sup>	.59	.36	.29	.39	.14	.35
F statistic	14.2	4.5	4.1	5.1	1.7	4.2
P value	<.001	<.001	<.001	<.001	.12	<.001

<sup>a</sup>Models exclude trials for which statistical significance could not be assessed (eg, safety trials).

<sup>b</sup>P < .05.

an endorsement by, the Office of the Assistant Secretary for Health, the HHS, or the US government.

## Open Science, Reproducibility, and Postpublication Peer Review

### Open Science Policies of Surgical Journals and the Use of Open Science Practices in Research Published in Surgical Journals

Jayson S. Marwaha,<sup>1,2</sup> Hao Wei Chen,<sup>1</sup> Harlan M. Krumholz,<sup>3</sup> Jeffrey B. Matthews<sup>4</sup>

**Objective** Reproducibility and transparency are important considerations in medical research<sup>1</sup>; recent retractions of studies in several medical journals underscore the relevance of these issues.<sup>2</sup> Many tools exist to promote research quality and transparency, including protocol preregistration sites for observational studies, EQUATOR Network reporting guidelines for most common study types, and preprint servers.<sup>3</sup> However, the extent to which the surgical research ecosystem has adopted these tools is unknown. The purpose of this study was to describe the use of these quality-promoting practices in surgical research.

**Design** Use of 5 open science practices were measured (preprint publication before peer-reviewed publication; use of EQUATOR Network guidelines; study protocol preregistration before peer-reviewed publication; published peer review; and public accessibility of data, experimental methods, and/or code) in surgical journals and manuscripts. A distinction was made between preregistration of clinical

trials in established trial registries (eg, ClinicalTrials.gov) and the emerging practice of preregistering outcomes and analysis plans for observational studies on newer platforms (eg, Protocols.io). The top 8 surgical journals by impact factor were included. A random sample of 240 research articles published from January 2019 to August 2021 by these journals (30 from each) were selected via random number generator and included in the study. The number of journals and studies that explicitly endorsed or used these practices was measured.

**Results** In their author guidelines, 7 of the 8 journals (88%) recommended the use of EQUATOR Network guidelines before journal submission. Five journals (63%) explicitly stated that they permitted submissions that were previously released as preprints. Only 3 journals (38%) recommended that authors preregister their protocols for observational studies. None published peer reviewer comments. Five (63%) explicitly recommended that authors make their methods, including all code, laboratory protocols, and data if possible, publicly available. Of 240 articles, 65 (27%) explicitly complied with the appropriate EQUATOR Network guideline. Only 30 observational studies (17%) preregistered their study protocols. None of the articles were posted on a preprint server before journal publication. Only 15 studies (6%) fully disclosed their methods in the form of making code public or publishing a separate protocol (**Table 21**). Research in the *International Journal of Surgery* exhibited the highest use of open science practices; studies in that journal used a mean of 1.9 open science practices vs 0.4 in the other journals ( $P < .001$ ). Journals that recommended (but did not explicitly require) compliance with an open science practice, such as

**Table 21. Adoption of Various Open Science Tools Currently Used to Promote Research Quality, Transparency, and Reproducibility<sup>a</sup>**

Open science tool	Journals, No. (%) (n = 8) <sup>b</sup>	Original research articles, No. (%) (n = 240)
EQUATOR Network guidelines	7 (88)	65 (27)
Preprints	5 (63)	0
Preregistration of clinical trials	7 (88)	30 (94)
Preregistration of observational studies	3 (38)	30 (17)
Published peer review	0	0
Open source methods	5 (63)	15 (6)

<sup>a</sup>Adoption of each tool was measured by checking if each journal's author guidelines explicitly endorsed its use and if each manuscript used the tool. For example, adoption of preprint servers was counted among journals if the author guidelines explicitly allowed for submission of preprint papers, and use of preprint servers was counted among original research articles if the article had been posted on a preprint server prior to peer-reviewed publication.

<sup>b</sup>JAMA Surgery, Annals of Surgery, British Journal of Surgery, Journal of the American College of Surgeons, Surgery, International Journal of Surgery, World Journal of Surgery, and American Journal of Surgery.

EQUATOR Network guideline use, had higher levels of open science practice use in their research vs journals that did not mention the practice at all (18% vs 4%;  $P < .001$ ). There was a positive association between journal impact factor and use of open science practices in its published studies ( $P < .001$ ).

**Conclusions** Surgical research is adapting slowly to open science practices in academia, leaving the field potentially vulnerable to poor research quality. There are many opportunities for improvement. The responsibility falls on both researchers and journals to consider a strategic model to adopt these new tools to promote high-quality research generation and dissemination.

## References

1. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
2. Ledford H, Van Noorden R. High-profile coronavirus retractions raise concerns about data oversight. *Nature*. 2020;582(7811):160.
3. PLOS. Transforming scientific communication through open science. Accessed January 26, 2022. <https://plos.org/open-science/>

<sup>1</sup>Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA, USA, jmarwaha@bidmc.harvard.edu; <sup>2</sup> Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA;

<sup>3</sup>Department of Medicine, Yale University School of Medicine, New Haven, CT, USA; <sup>4</sup>Department of Surgery, University of Chicago Pritzker School of Medicine, Chicago, IL, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** Jayson S. Marwaha is supported by a grant from the National Library of Medicine/National Institutes of Health (T15LM007092-30) and the Biomedical Informatics and Data Science Research Training (BIRT) Program of Harvard University.

**Additional Information** Jayson S. Marwaha and Hao Wei Chen are co-first authors and Harlan M. Krumholz and Jeffrey B. Matthews are co-senior authors.

## Characteristics of Studies of Research Reproducibility in Economics, Education, Psychology, Health Sciences, and Biomedicine: A Scoping Review

Kelly D. Cobey,<sup>1,2</sup> Christophe A. Fehlmann,<sup>2,3</sup> Marina Christ Franco,<sup>4,5</sup> Ana Patricia Ayala,<sup>6</sup> Lindsey Sikora,<sup>7</sup> Danielle B. Rice,<sup>4,8</sup> Chenchen Xu,<sup>4,9</sup> John P. A. Ioannidis,<sup>10</sup> Manoj M. Lalu,<sup>4,11,12</sup> Alixe Ménard,<sup>4</sup> Andrew Neitzel,<sup>4,9</sup> Bea Nguyen,<sup>4,9</sup> Nino Tsartsadze,<sup>4</sup> David Moher<sup>4,2</sup>

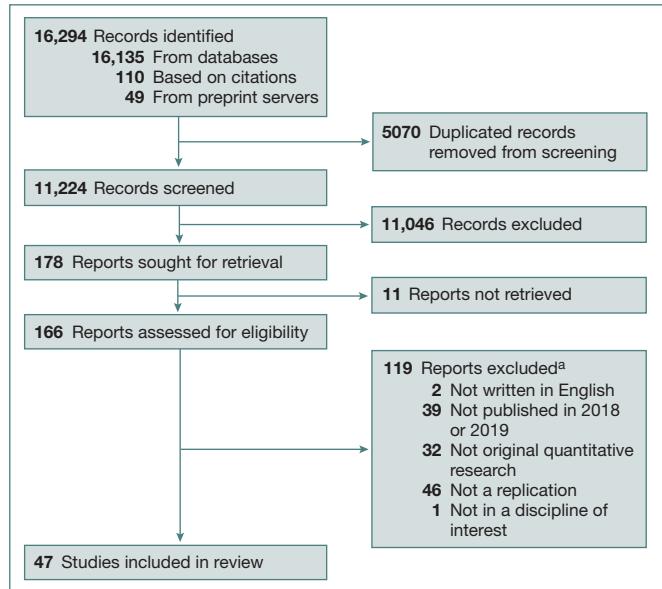
**Objective** Reproducibility is a central tenet of research. Explicit reproducibility checks are made across different disciplines trying to assess the replicability of previously published studies. This scoping review aimed to synthesize the literature on reproducibility; describe its epidemiological characteristics, including how reproducibility is defined and assessed; and determine and compare estimates for reproducibility across different fields.

**Design** All English-language quantitative replication studies within the fields of economics, education, psychology, health sciences, and biomedicine published in 2018 or 2019 were included, as were studies that were explicitly self-described as a replication or a reproducibility study in which a previously published quantitative study is referred to and conducted again. Conference proceedings, commentaries, narrative reviews, systematic reviews, and clinical case studies were excluded. MEDLINE, Embase, PsycINFO, Cumulative Index of Nursing and Allied Health Literature (CINAHL), Education Source via EBSCOHost, ERIC, EconPapers, International Bibliography of the Social Sciences, and EconLit were searched. Retrieved documents were screened in duplicate against our inclusion criteria. The year of publication, number of authors, country of affiliation of the corresponding author, and funding were extracted. For individual replication studies, whether a registered protocol was used, whether there was contact between the replicating team and the original authors, what study design was used, and what the primary outcome was were extracted from each replication study. Finally, how replication was defined by the authors and whether the assessed study or studies successfully replicated a previous study based on this definition were also extracted. Extraction was done by 2 reviewers.

**Results** The search identified 11,224 unique documents, of which 47 were included (Figure 14). Most studies were related to either psychology (48.6%) or health sciences (23.7%). Among these 47 documents, 36 described a single replication study, while the remaining 11 reported at least 2 replications in the same paper. Less than half of the studies referred to a registered protocol. There was variability in the definitions of replication success, with studies related to psychology and health sciences tending to use a comparison of effect sizes to define replication success. Based on the definition used by the authors of each study, 95 of 177 studies (53.7%) achieved replication success.

**Conclusions** This study gives an overview of research across 5 disciplines that explicitly set out to replicate previous

**Figure 14. Study Flow Diagram**



research. While estimates of reproducibility vary across fields in this modest sample, so too do norms in definitions used to define reproducibility.

<sup>1</sup>University of Ottawa Heart Institute, Ottawa, ON, Canada, kcobey@ottawaheart.ca; <sup>2</sup>School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada; <sup>3</sup>Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals, Geneva, Switzerland; <sup>4</sup>Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada; <sup>5</sup>School of Dentistry, Federal University of Pelotas, Pelotas, Brazil; <sup>6</sup>Gerstein Science Information Centre, University of Toronto, Toronto, ON, Canada; <sup>7</sup>Health Sciences Library, University of Ottawa, Ottawa, ON, Canada; <sup>8</sup>Department of Psychology, McGill University, Montreal, QC, Canada; <sup>9</sup>Department of Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>10</sup>Departments of Medicine, of Epidemiology and Population Health, of Biomedical Data Science, and of Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA; <sup>11</sup>Department of Anesthesiology and Pain Medicine, The Ottawa Hospital, University of Ottawa, Ottawa, ON, Canada; <sup>12</sup>Regenerative Medicine Program, The Ottawa Hospital, Ottawa, ON, Canada; <sup>13</sup>Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada

**Conflict of Interest Disclosures** David Moher is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision of this abstract. No other disclosures were reported.

**Additional Information** Alixe Ménard is a patient author.

### Data Sharing and Reanalysis for Main Studies Assessed by the European Medicines Agency

Maximilian Siebert,<sup>1,2</sup> Jeanne Gaba,<sup>1,2</sup> Alain Renault,<sup>1,2</sup> Bruno Laviolle,<sup>1,2</sup> Clara Locher,<sup>1,2</sup> David Moher,<sup>3</sup> Florian Naudet<sup>1,2</sup>

**Objective** Transparency and reproducibility are expected to be normative practices in clinical trials used for decision-making on marketing authorizations for new medicines. A cross-sectional study was conducted aiming to assess

inferential reproducibility for main trials (sometimes referred to as pivotal trials) assessed by the European Medicines Agency.

**Design** Two members of the team (J.G., M.S.) independently identified all studies on new medicines, biosimilars, and orphan medicines given approval by the European Commission between January 2017 and December 2019, categorized as main studies in the European Public Assessment Reports (EPARs). Sixty-two of 292 eligible studies were randomly sampled. One team member (J.G.) identified the sponsors and sent a standardized message to retrieve the individual patient data (IPD) for these studies. Up to 3 reminder messages were sent. A dossier for each study was prepared containing the IPD, the protocol, and information on the conduct of the study. A second team member (M.S.), who had no access to the study reports, used the dossier to run an independent reanalysis of each trial. All results of these reanalyses were reported in terms of each study's conclusions, *P* values, effect sizes, and changes from the initial protocol. Two team members (J.G., F.N.) not involved in the reanalysis compared the results of the reanalyses with the published results of the trial.

**Results** A total of 292 main studies in 173 EPARs were identified. Among the 62 studies randomly sampled, IPD was received for 10 trials (16%). The median number of days between data request and data receipt was 253 (IQR, 182–469). For these 10 trials, 23 distinct primary outcomes were identified for which the conclusions were reproduced in all reanalyses. Therefore, 10 of 62 trials (16%; 95% CI, 8%–28%) were reproduced. Regarding the 52 studies without available data, assessment of reproducibility was not possible. Forty-eight of the 52 sponsors replied to the request. The reasons for nonsharing can be found in Table 22. There was no change from the original study protocol regarding the primary outcome in any of the 10 studies. Spin (ie, interpretation bias) was observed in the report of 1 study.

**Conclusions** Despite their results supporting decisions that affect millions of people's health across the European Union, most of the main studies used in EPARs lack transparency as data were not shared with external researchers to assess reproducibility. The limits of this approach lie in the small amount of IPD obtained. Nonetheless, reanalyses of the few trials with available data showed complete inferential reproducibility. Further studies with a larger sample size are necessary to estimate the reproducibility of clinical trials included in the marketing authorizations.

<sup>1</sup>University of Rennes, INSERM, Centre d'Investigation Clinique de Rennes, Rennes, France, maximilian.siebert@univ-rennes1.fr; <sup>2</sup>University of Rennes, INSERM, Institut de Recherche en Santé, Environnement et Travail, Rennes, France; <sup>3</sup>Center for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

**Conflict of Interest Disclosures** David Moher is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. No other disclosures were reported.

**Table 22. Reasons for Not Sharing Individual Patient Data (N = 52)**

Restriction owing to study status (n = 13)
Confidentiality (n = 9)
Doubt of scientific merit (n = 9)
No explicit reason (n = 5)
No response (n = 4)
Complexity of the original study (n = 3)
No data sharing initiative (n = 3)
Data sharing policy does not extend to data sharing (n = 2)
Inability to request data (n = 2)
Patient informed consent (n = 1)
Risk of reidentification of patients (n = 1)

**Funding/Support** The project is funded by the Agence Nationale de la Recherche.

**Role of the Funder/Sponsor** The sponsor had no role design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

### Assessment of Postpublication Critique Policies and Practices at Top-Ranked Journals in 22 Scientific Disciplines

Tom E. Hardwicke,<sup>1</sup> Robert T. Thibault,<sup>2,3</sup> Jessica E. Kosie,<sup>4</sup> Loukia Tzavella,<sup>5</sup> Theiss Bendixen,<sup>6</sup> Sarah A. Handcock,<sup>7</sup> Vivian E. Köneke,<sup>7</sup> John P. A. Ioannidis<sup>2,8,9</sup>

**Objective** To describe how top-ranked journals across 22 scientific disciplines handle postpublication critique such as letters, commentaries, and online comments.<sup>1-3</sup>

**Design** Cross-sectional assessment of policies and practices related to postpublication critique at 15 journals (top-ranked by impact factor) operating in each of 22 scientific disciplines (defined by Clarivate Essential Science Indicators) assigned to 5 high-level scientific domains (defined by the authors; 330 journals). Policy information was extracted from journal websites in November 2019. For each journal offering postpublication critique, a random sample of 10 research articles published in 2018 (2066 articles) was examined to see if they were linked to postpublication critique on the article's webpage (1 journal only published 6 articles in 2018). Features of all linked postpublication critiques and associated author replies were recorded.

**Results** Overall, 207 of 330 journals (63%) offered postpublication critique such as letters (118), commentaries (85), or web comments (41) but often imposed limits on length (median, 1000; IQR, 500-1200 words) and time to-submit (median, 12; IQR, 4-26 weeks). The most restrictive limits were 175 words and 2 weeks; the least restrictive policies had no limits. Seventy-four journal policies implied independent external peer review of postpublication critique. Of a random sample of 2066 research articles published by journals offering postpublication critique, 39 (1.9%; 95% CI, 1.4%-2.6%) were linked to at least 1 postpublication critique

(there were 58 postpublication critiques in total). Of these target articles, 34 were from the health and life sciences and 5 were from multidisciplinary journals. Examination of all 58 postpublication critiques found that they addressed issues related to design (19), implementation (3), analysis (19), reporting (10), interpretation (45), and ethics (1); 29 were paywalled; 45 had conflict of interest statements, 15 of which declared a potential conflict; 44 received an author reply, of which 41 asserted that the authors' conclusions were unchanged. Fifty-one did not include any novel statistical analyses of original or new data, though only 3 target articles stated that data were available. The health and life sciences and multidisciplinary journals offered and published more postpublication critiques relative to other domains (**Table 23**). Clinical medicine in particular stood out, with the highest prevalence of postpublication critique (13% of 150 articles) and all 15 journals allowing postpublication critique. However, these journals also imposed the strictest limits on length (median, 400; IQR, 400-550 words) and time to submit (median, 4; IQR, 4-6 weeks).

**Conclusions** Top-ranked academic journals across scientific disciplines often pose barriers to the cultivation, documentation, and dissemination of postpublication critique. Publication of postpublication critique was rare in most disciplines. Published postpublication critique may have little effect on authors' conclusions.

### References

- Bastian H. A stronger post-publication culture is needed for better science. *PLoS Medicine*. 2014;11:e1001772. doi:10.1371/journal.pmed.1001772
- Altman DG. Poor-quality medical research: what can journals do? *JAMA*. 2002;287:2765-2767. doi:10.1001/jama.287.21.2765
- Winker MA. Letters and comments published in response to research: whither postpublication peer review? Abstract presented at: International Congress on Peer Review and Scientific Publication; September 9, 2013; Chicago, Illinois. <https://peerreviewcongress.org/2013-abstracts/>

<sup>1</sup>Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; <sup>2</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA; <sup>3</sup>School of Psychological Science, University of Bristol, Bristol, UK;

<sup>4</sup>Department of Psychology, Princeton University, Princeton, NJ, USA; <sup>5</sup>School of Psychology, Cardiff University, Cardiff, UK; <sup>6</sup>Department of the Study of Religion, Aarhus University, Aarhus, Denmark; <sup>7</sup>Florey Department of Neuroscience and Mental Health, University of Melbourne, Melbourne, Australia;

<sup>8</sup>Charité–Universitätsmedizin Berlin, Berlin, Germany; <sup>9</sup>Meta-Research Innovation Center Berlin (METRIC-B), QUEST Center for Transforming Biomedical Research, Berlin Institute of Health, Charité–Universitätsmedizin Berlin, Berlin, Germany;

<sup>10</sup>Departments of Medicine, Epidemiology and Population Health, Biomedical Data Science, and Statistics, Stanford University, Stanford, CA, USA

**Conflict of Interest Disclosures** Tom E. Hardwicke receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant

**Table 23. Key Features of Postpublication Critique Policy and Practice at 330 Top-Ranked Journals From 22 Scientific Disciplines Stratified Into 5 High-Level Domains**

Feature	All domains (330 journals) <sup>a</sup>	Health and life sciences (165 journals)	Multidisciplinary (15 journals)	Physical sciences (90 journals)	Formal sciences (30 journals)	Social sciences (30 journals)
Journals offering any PPC, No. (%) <sup>b</sup>	207 (63)	126 (76)	11 (73)	53 (59)	9 (30)	8 (27)
Journals offering letters, No. (%)	118 (36)	84 (51)	5 (33)	21 (23)	5 (17)	3 (10)
Journals offering commentaries, No. (%)	85 (26)	38 (23)	6 (40)	33 (37)	4 (13)	4 (13)
Journals offering web comments, No. (%)	41 (12)	32 (19)	7 (47)	2 (2)	0	0
Journals imposing length limits, No.	Any: 167 quant: 162	Any: 115 quant: 111	Any: 11 quant: 10	Any: 26 quant: 26	Any: 9 quant: 9	Any: 6 quant: 6
Quantitative length limits, median (IQR), words	1000 (500-1200)	750 (500-1200)	1000 (500-1150)	1200 (1200-1500)	1000 (1000-1500)	3000 (1500-8250)
Journals imposing time-to-submit limits, No.	Any: 80 quant: 49	Any: 64 quant: 37	Any: 4 quant: 3	Any: 10 quant: 7	Any: 1 quant: 1	Any: 1 quant: 1
Quantitative time-to-submit limits, median (IQR), weeks	12 (4-26)	8 (4-17)	13 (13-20)	52 (26-78)	52	26
Articles with PPC, No. (% [95% CI])	39/2066 (1.9 [1.4-2.6])	34/1256 (2.7 [1.9-3.8])	5/110 (4.5 [1.7-10.8])	0/530 (0 [0-0.9])	0/90 (0 [0-5.1])	0/80 (0 [0-5.7])

Abbreviations: PPC, postpublication critique; quant, quantitative.

<sup>a</sup>Health and life sciences included pharmacology and toxicology, clinical medicine, immunology, microbiology, biology and biochemistry, plant and animal science, environment and ecology, agricultural sciences, neuroscience and behavior, molecular biology and genetics, and psychiatry and psychology. Multidisciplinary included multidisciplinary journals: physical sciences, space science, physics, chemistry, geosciences, engineering, materials science. Formal sciences included mathematics, computer science. Social sciences included general social sciences, economics and business.

<sup>b</sup>Some journals offered multiple types of postpublication critique. Consequently, the number of journals offering letters, commentaries, or web comments does not sum to the number of journals offering any postpublication critique.

agreement No. 841188. Robert T. Thibault is supported by a general support grant awarded to METRICS from the Laura and John Arnold Foundation and a postdoctoral fellowship from the Fonds de recherche du Québec—Santé. Theiss Bendixen thanks the Aarhus University Research Foundation for support. Jessica E. Kosie received funding from NSF SBE Postdoctoral Research Fellowship 2004983 and NIH F32 National Research Service Award HD103439. Loukia Tzavella was supported by ESRC postdoctoral fellowship ES/Vo11030/1. No other conflicts were reported.

**Funding/Support** The Meta-Research Innovation Center at Stanford (METRICS) is supported by a grant from the Laura and John Arnold Foundation. The Meta-Research Innovation Center Berlin (METRIC-B) is supported by a grant from the Einstein Foundation and Stiftung Charité.

**Role of the Funder/Sponsor** The funders had no role in this research.

## Social Media and Citations

### Association of Medical Research Visual Abstract Display With Social Media–Driven Site Traffic

N. Seth Trueger,<sup>1,2</sup> Eman Aly,<sup>3</sup> Sébastien Haneuse,<sup>2,4</sup> Evelyn Huang,<sup>1</sup> Michael Berkwits<sup>3</sup>

**Objective** Medical journals often use visual abstracts (VAs), “infographics” designed to graphically convey a study’s research question, methods, findings, and conclusions, to summarize and promote published research on social media. Studies suggest that VAs increase an article’s reach and engagement, mostly demonstrated for individual journals on Twitter.<sup>1-3</sup> The JAMA Network, a family of 2 general medical and 11 specialty journals, began creating VAs for randomized clinical trials (RCTs) in 2018-2020 and publishing them online and in social media. This study examined the

association of VAs with social media–driven site traffic and metrics.

**Design** This random-sequence simultaneous crossover trial included all RCTs published in the 13 JAMA Network journals with an accompanying VA between September 21, 2021, and May 15, 2022. For each, tweets and Facebook posts promoting the RCT were published including a standard text summary with a link to the online article, accompanied by 3 image display types: (1) preview VA, a VA thumbnail that linked to the article when clicked; (2) expandable VA, a VA thumbnail that expanded to full screen when clicked; and (3) article image, an article figure or table thumbnail that linked to the article (Facebook) or expanded to full screen (Twitter) when clicked (control display). Three tweets and 3 Facebook posts, 1 with each image display type, were published to Twitter and Facebook nearly simultaneously (1 minute apart) in random sequence within an hour of article publication. The primary outcome was the number of link clicks by image display type at 1 week. Secondary outcomes were social media Twitter impressions/Facebook reach (how many times people saw the tweet or post) and engagement (sum total number of comments/replies, shares by other users, and likes) by image display type at 1 week. The study was designed to provide 80% power to detect a 50% increase in median link clicks comparing either of the 2 VA formats vs article image display.

**Results** Among 205 JAMA Network RCTs with VAs, link clicks to full-text articles were higher with preview VAs vs article images; impressions/reach and engagement were higher with expandable VAs vs article image displays (**Table 24**). In a preplanned subgroup analysis, the higher link clicks appeared attributable to preview VA display on Twitter (preview VA median, 12 [IQR, 3-41] and expandable VA median, 6 [IQR, 2-21] vs article image median, 6 [IQR,

**Table 24. Social Media–Driven Site Traffic, Reach, and Engagement by Image Display Type**

	Median (IQR)			P value <sup>a</sup>
	Preview VA	Expandable VA	Article image	
Link clicks <sup>b</sup>	18 (4-64)	11 (3-38)	9 (2-40)	.02
Reach/ impressions <sup>c</sup>	1874 (685-6874)	3344 (1102-17,008)	2082 (765-7957)	<.001
Engagement <sup>d</sup>	24 (7-76)	43 (12-162)	21 (6-77)	.001

Abbreviation: VA, visual abstract.

<sup>a</sup>By Kruskal-Wallis test.

<sup>b</sup>Primary outcome.

<sup>c</sup>Reach (Facebook)/impressions (Twitter): how many times people saw the post.

<sup>d</sup>Engagement: sum of user comments/replies, shares by other users, and likes.

1-21];  $P < .001$ ) rather than Facebook (medians, 1 [IQR, 0-14] and 3 [IQR, 0-12] vs 1 [IQR, 0-13], respectively;  $P = .11$ ).

**Conclusions** In this random-sequence simultaneous crossover trial, the use of visual abstracts in social media posts was associated with higher social media–driven journal site traffic when displayed as preview links, and higher social media reach and engagement when displayed as expandable images, compared with article images.

## References

1. Oska S, Lerma E, Topf J. A picture is worth a thousand views: a triple crossover trial of visual abstracts to examine their impact on research dissemination. *J Med Internet Res*. 2020;22(12):e22327. doi:10.2196/22327
2. Ibrahim AM, Lillemoe KD, Klingensmith ME, Dimick JB. Visual abstracts to disseminate research on social media: a prospective, case-control crossover study. *Ann Surg*. 2017;266(6):e46-e48. doi:10.1097/SLA.oooooooooooooo2277
3. Chisari E, Gouda Z, Abdelaal M, et al. A crossover randomized trial of visual abstracts versus plain-text tweets for disseminating orthopedics research. *J Arthroplasty*. 2021;36(8):3010-3014. doi:10.1016/j.arth.2021.04.006

<sup>a</sup>Department of Emergency Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, ntrueger@gmail.com; <sup>b</sup>JAMA Network Open, Chicago, IL, USA; <sup>c</sup>JAMA Network, Chicago, IL, USA; <sup>d</sup>Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA

**Conflict of Interest Disclosures** N. Seth Trueger receives salary support from the American Medical Association (AMA) for his role as digital media editor, *JAMA Network Open*. Eman Aly and Michael Berkwits are full-time employees of the AMA and the JAMA Network. Sebastien Haneuse receives compensation from the AMA for his role as statistical editor, *JAMA Network Open*. Michael Berkwits is co-director of the Peer Review Congress but was not involved in the review or decision for this abstract.

## Evaluation of Editors' Abilities to Estimate Citation Potential of Research Manuscripts Submitted to *The BMJ*

Sara Schroter,<sup>1</sup> Wim Weber,<sup>1</sup> Elizabeth Loder,<sup>1</sup> Jack Wilkinson,<sup>2</sup> Jamie J. Kirkham<sup>2</sup>

**Objective** To evaluate editors' ability to estimate the citation potential of a cohort of research submissions after publication.

**Design** Research manuscripts submitted to *The BMJ*, sent for peer review, and subsequently scheduled for discussion at an editorial meeting between August 27, 2015, and December 29, 2016, were rated independently by attending editors for citation potential prior to discussion at meetings. For each manuscript, editors indicated how many citations they thought each manuscript would generate in the year they were first published plus the first calendar year after publication, in relation to the median number of citations for a paper published in *The BMJ* at the time. Editors could choose from the following 4 categories: no citations; below *The BMJ* average number of citations (<10); around *The BMJ* average number of citations (10-17); and more than *The BMJ* average number of citations (>17). Google, PubMed, ResearchGate, institutional websites, ORCID, and trial registries were searched for subsequent journal publications using key information submitted by authors. Actual citations generated were extracted from the Web of Science (WOS) Core Collection on May 10, 2022. To ensure citation counts were complete, analysis was restricted to articles published by December 31, 2019, or not published at the time of analysis.

**Results** Of 530 manuscripts, 508 were published as full-length articles and indexed in the WOS and 22 were unpublished (1 abstract, 1 preprint, 1 substantially changed, 19 not found). Among the 507 manuscripts published by the end of 2019, the median (IQR [range]) number of citations in the year of publication plus the following year was 8 (4-16 [0-150]). A total of 291 manuscripts (57%) generated below *The BMJ* average number of citations (<10), 102 (20%) generated around *The BMJ* average number of citations (10-17), and 114 (23%) generated above *The BMJ* average number of citations (>17). The number of citations was higher for accepted manuscripts (median, 12 [IQR, 7-24] citations) compared with rejected manuscripts (median, 5 [IQR, 3-10.75] citations). For each of the 10 editors' ratings, there was a tendency for actual citation counts to be higher in line with the editor's increasing estimated citation categories but with considerable variation within categories; 9 of 10 editors were unable to identify the correct citation category for more than 50% (range, 31%-54%) of manuscripts. A  $\kappa$  analysis revealed that agreement between the estimated and actual categories for all editors was slight or fair ( $\kappa$  value range, 0.02-0.27). **Table 25** shows that editors frequently rated papers that were highly cited as having low citation potential and vice versa. Secondary analysis using citations in the first 2 years after publication showed similar results.

**Conclusions** Many editors are motivated to publish highly citable manuscripts because this determines impact factor; however, this motivation can bias which articles get published and where they are published. This study found that *The BMJ* editors were not good at estimating the citation potential of manuscripts they accepted or rejected.

**Table 25. Analysis of the Extreme Disagreements for Participating Editors**

Editor	No. of papers with low citations rated <sup>a</sup>	Papers with low citations rated as having high citation potential, No. (%) <sup>a</sup>	No. of highly cited papers rated <sup>b</sup>	Highly cited papers rated as having low citation potential, No. (%) <sup>b</sup>
A	215	3 (1)	75	38 (51)
B	44	5 (11)	12	6 (50)
C	285	79 (28)	112	26 (23)
D	148	27 (18)	60	20 (33)
E	218	23 (11)	97	31 (32)
F	232	2 (1)	85	51 (60)
G	49	11 (22)	26	10 (38)
H	121	5 (4)	45	18 (40)
I	30	3 (10)	11	5 (45)
J	22	4 (18)	3	1 (33)

<sup>a</sup>Manuscripts that generated fewer than 10 citations in the Web of Science Core Collection in the year of publication plus the following year.

<sup>b</sup>Manuscripts that generated more than 17 citations in the Web of Science Core Collection in the year of publication plus the following year.

<sup>1</sup>The BMJ, London, UK, sschroter@bmj.com; <sup>2</sup>The University of Manchester, Manchester, UK

**Conflict of Interest Disclosures** Sara Schroter, Wim Weber, and Elizabeth Loder are employed by or seconded to *The BMJ*. Jack Wilkinson holds statistical or methodologic editor roles for Cochrane Gynaecology and Fertility, *BJOG: An International Journal of Obstetrics and Gynaecology, Reproduction and Fertility*, and *Fertility and Sterility*. Jamie J. Kirkham is a statistical editor for *The BMJ*.

**Acknowledgments** Thanks to Nillanee Nehrujee for assistance with data collection for articles accepted to *The BMJ* and to *The BMJ* research editors for their participation.

**Additional Information** Jack Wilkinson is a co-corresponding author.

### Improper Legitimization of Hijacked Journals Through Citations

Anna Abalkina,<sup>1</sup> Guillaume Cabanac,<sup>2</sup> Cyril Labbé,<sup>3</sup> Alexander Magazinov<sup>4</sup>

**Objective** Hijacked journals are fake websites that use the titles, International Standard Serial Numbers (ISSNs), and metadata of established journals to appear legitimate so they can collect publishing fees from unsuspecting authors without peer review. There are more than 200 documented cases of hijacked journals,<sup>1</sup> most of which publish low-quality or dishonest research. Moussa documented citations in legitimate journals to hijacked journal articles in the field of marketing.<sup>2</sup> This study sought to document the number of citations in legitimate journals to articles in hijacked journals, thus potentially legitimizing those unreliable articles and fraudulent journals.

**Design** In this cross-sectional study, a citejacked detector, designed as a part of the Problematic Paper Screener,<sup>3</sup> articles in legitimate journals citing hijacked journal articles (ie, citejacked articles) were tabulated. The detector screened 12

journals documented to be hijacked, with the hijacked versions erroneously indexed in international bibliographic databases. A full-text search was performed between November 2021 and January 2022 with Dimensions, a bibliometric database containing more than 100 million publications, using the name of 1 of the 12 hijacked journals and including articles published and indexed between January 1, 2021, and January 31, 2022. To exclude items published in preprints or predatory journals, the search was limited to the articles published in journals listed in the Norwegian Register for Scientific Journals, Series and Publishers. Each retrieved item was manually checked to retain true positives (citations to hijacked journals) and discard false positives (citations to authentic journals or other mentions of journals).

**Results** Of 1421 articles featuring the title of a hijacked journal, 828 (58.3%) cited unreliable articles from hijacked journals. Citejacked articles were published by 67 publishers.

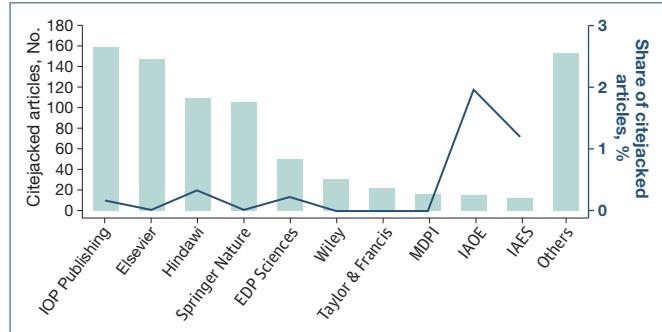
**Figure 15** shows the distribution of citejacked articles among the top 10 publishers. Flagship publishers were not immune to incorporating references to hijacked journals in their citation index. Between January 1, 2021, and January 31, 2022, a mean of 2 citejacked articles were published daily in established journals.

**Conclusions** Reputable journals cite unreliable articles from hijacked journals, legitimizing such pseudo-articles. The presence of citejacked articles in reference lists points to a flaw in the peer review process and shows that curation of references must be taken more seriously. These flawed references to illegitimate journals may serve as markers of articles that are problematic due to plagiarism in articles originating from hijacked journals, citation cartels, and paper mills. Given the limited number of titles included in this study (12 among the more than 200 documented<sup>2</sup>), the phenomenon might be wider and has not yet been systematically studied.

### References

1. Abalkina A. Detecting a network of hijacked journals by its archive. *Scientometrics*. 2021;126:7123-7148. doi:10.1007/s11192-021-04056-0

**Figure 15. Number and Share of Citejacked Articles per Publisher Between January 2021 and January 2022**



Distribution of 828 citejacked articles among the top 10 publishers. IAEO indicates International Association of Online Engineering; IAES, Institute of Advanced Engineering and Science; and MDPI, Multidisciplinary Digital Publishing Institute.

2. Moussa S. A “Trojan horse” in the reference lists: citations to a hijacked journal in SSCI-indexed marketing journals. *J Acad Librarianship*. 2021;47(5). doi:10.1016/j.acalib.2021.102388

3. Problematic Paper Screener. February 27, 2021. <https://www.irit.fr/~Guillaume.Cabanac/problematic-paper-screener>

<sup>1</sup>Freie Universität Berlin, Berlin, Germany, anna.abalkina@fu-berlin.de; <sup>2</sup>Computer Science Department, University of Toulouse, Institut de Recherche en Informatique de Toulouse, Toulouse, France; <sup>3</sup>Université Grenoble Alpes, National Centre for Scientific Research, Grenoble, France; <sup>4</sup>Yandex, Moscow, Russia

**Conflict of Interest Disclosures** None reported.

**Funding/Support** Cyril Labb  and Guillaume Cabanac were supported by ERC-Synergy NanoBubbles (ERC-2020-SyG-951393)

# Poster Session Abstracts

**Posters will be presented during 1 of 2 sessions, on Friday, September 9, and Saturday, September 10 and will be available online throughout the meeting.**

**Virtual Posters will be available online September 8-10.**

**All posters and related materials will be available online after the meeting.**

**FRIDAY, SEPTEMBER 9**

## Artificial Intelligence

### Counterfactual Evaluation of Peer Review Assignment Strategies in Computer Science and Artificial Intelligence

Martin Saveski,<sup>1</sup> Steven Jecmen,<sup>2</sup> Nihar Shah,<sup>2</sup> Johan Ugander<sup>1</sup>

**Objective** Artificial intelligence (AI) has become pervasive to assign reviewers to papers.<sup>1</sup> The assignment relies on 3 key sources of data<sup>1</sup>: (1) AI-computed similarities between the text of the submitted paper and reviewers' past articles, (2) reviewer-provided preferences expressing which papers they would like to review, and (3) overlap between the paper's topics as specified by authors and reviewers' self-reported areas of expertise. However, it is unknown which of these sources, or combination thereof, lead to the best outcomes of the reviewer assignment.

**Design** To assign reviewers to papers, 2 venues recently used randomized algorithms<sup>2</sup> designed to combat fraud: the 2021 Theory and Practice of Differential Privacy (TPDP) Workshop with 35 reviewers and 95 full papers and the Association for the Advancement of Artificial Intelligence (AAAI) 2022 Conference on Advancement in Artificial Intelligence with 3145 reviewers and 8450 full papers. To compute overall similarities between each reviewer-paper pair, TPDP weighted the AI-computed text similarities by weight ( $w_{text}$ , range 0-1) and reviewers' preferences by weight ( $1 - w_{text}$ ); AAAI weighted the AI-computed text similarities by weight ( $w_{text}$ , range 0-1) and the overlap between the papers and reviewers' topical areas by weight ( $1 - w_{text}$ ) (reviewers' preferences were also included in AAAI but not considered in this study). The randomized assignment<sup>2</sup> then maximized similarity of assigned reviewer-paper pairs, subject to the probability of any reviewer being assigned to any paper being at most 0.5 in TPDP and 0.52 in AAAI. In this study, the randomization in the assignment was leveraged to estimate the counterfactual quality of alternative assignment strategies. How the overall quality of the reviewer-paper assignment was affected was investigated by (1) introducing randomness in the assignment process and (2) varying weights of different sources of information. The quality of any counterfactual reviewer-paper assignments was measured

using reviewers' self-reported expertise and confidence in their review.

**Results** The results are tabulated in **Table 26**.<sup>3</sup> First, introducing randomness by limiting the probability of any reviewer-paper assignment led to a marginal reduction in assignment quality for TPDP and a slightly larger reduction in AAAI. Second, for TPDP, placing more weight on the AI-computed text similarities ( $w_{text} = 0.8$ ) instead of equally weighting the text similarities and the reviewers' preferences ( $w_{text} = 0.5$ ) resulted in a higher reviewer-paper assignment quality. Third, for AAAI, placing more weight on the AI-computed text similarities ( $w_{text} = 0.75$ ) instead of equally weighting the text similarity and the reviewer-paper topical area overlap ( $w_{text} = 0.5$ ) led to a similar assignment quality.

**Conclusions** Randomness in the reviewer assignments can help improve AI-based automated assignment by enabling counterfactual analysis of alternative assignment strategies, in addition to its original goal of mitigating fraud, but leads to a small reduction in assignment quality.

### References

- Shah N. Challenges, experiments, and computational solutions in peer review. *Commun ACM*. 2022;65(6):76-87. doi:10.1145/3528086
- Jecmen S, Zhang H, Liu R, Shah N, Conitzer V, Fang F. Mitigating manipulation in peer review via randomized reviewer assignments. *Adv Neural Inf Process Syst*. 2020;33:12533-12545.

**Table 26. Mean Reviewer Expertise and Confidence for Different Reviewer-Paper Assignment Strategies**

Venue	Strategy	$w_{text}$	Expertise <sup>a,b</sup>	Confidence <sup>a,b</sup>
TPDP	Randomized	0.5	2.63 (2.61-2.65)	3.41 (3.37-3.43)
TPDP	Deterministic	0.5	2.68 (2.63-2.72)	3.42 (3.35-3.47)
TPDP	Deterministic	0.8	2.78 (2.73-2.83)	3.61 (3.53-3.66)
AAAI	Randomized	0.75	3.43 (3.14-3.56)	2.88 (2.64-2.97)
AAAI	Deterministic	0.75	3.68 (3.43-3.77)	3.06 (2.85-3.12)
AAAI	Deterministic	0.5	3.66 (2.77-3.99)	3.06 (2.31-3.29)

Abbreviations: AAAI, Association for the Advancement of Artificial Intelligence; TPDP, Theory and Practice of Differential Privacy.

<sup>a</sup>The numbers in the parentheses show 95% CIs for the TPDP Workshop and Manski bounds<sup>3</sup> around the mean for the AAAI Conference on Advancement in Artificial Intelligence; both intervals account for reviewer attrition.

<sup>b</sup>For TPDP, expertise scores ranged between 1 (irrelevant) and 4 (very relevant), and confidence scores ranged between 1 (educated guess) and 5 (absolutely certain). For AAAI, expertise scores ranged between 0 (not knowledgeable) and 5 (expert), and confidence scores ranged between 0 (not confident) and 4 (very confident).

3. Imbens GW, Manski CF. Confidence intervals for partially identified parameters. *Econometrica*. 2004;72(6):1845-1857. doi:10.1111/j.1468-0262.2004.00555.x

<sup>1</sup>Stanford University, Stanford, CA, USA; <sup>2</sup>Carnegie Mellon University, Pittsburgh, PA, USA, nihars@cs.cmu.edu

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported by the US National Science Foundation CAREER award (1942124) which supports research on the fundamentals of learning from people with applications to peer review.

**Acknowledgments** We thank Gautam Kamath and Rachel Cummings for allowing us to conduct this study in TPDP and Melisa Bok and Celeste Martinez Gomez from OpenReview.net for helping with the APIs of OpenReview.net.

### Utility of Machine Learning in Predicting Success of a Peer Review Paper From Peer Reviewer Scores

Ernest Kimani,<sup>1</sup> James Kigera,<sup>2</sup> Vincent Kipkorir<sup>1</sup>

**Objective** To investigate the utility of machine learning algorithms in predicting the likelihood of publication of a manuscript from peer reviewer scores.

**Design** Using a cross-sectional study design, 263 manuscripts that had undergone peer review between 2017 and 2021 were selected, and a final decision made on whether to accept or reject a manuscript for publication. Excluded were manuscripts with incomplete data on peer reviewer scores. Data were collected on the manuscripts' peer reviewer scores and the final decision made by the journal. Peer reviewers' scores included ratings by 2 peer reviewers per manuscript on originality, quality, interest, overall rating, and priority for publishing. Two-thirds (174 manuscripts) of the data were used for training and one-third (89 manuscripts) for testing the algorithms. Microsoft Excel 2019 was used to preprocess the data and Weka version 3.9.5 was used for model assessment. Training and testing of the model were conducted using various machine learning algorithms. The model with the highest accuracy in predicting the likelihood of a manuscript to be published would be further improved and deployed for application.

**Results** One-hundred and thirty-four manuscripts were accepted for publication and 129 rejected for the final analysis. The performance of various machine learning algorithms in predicting the likelihood of publication ranged from 58.4% to 65.2% (**Table 27**).

**Conclusions** A machine learning model to reduce peer review workload would ensure that scarce peer review resources are utilized by optimizing desk rejections. Such a model would promote efficiency in the publishing process and improve overall journal output and satisfaction with authors. To implement such a model, the in-house reviewers and editorial team could score manuscripts and assess their performance before advancing them for peer review. To improve the model's performance and reduce bias, there

**Table 27. Accuracy of Machine Learning (ML) Algorithms in Predicting the Likelihood of Manuscript Publication**

ML classifier	Testing manuscripts, No. (n = 89)	Accuracy, %
Naive Bayes	58	65.2
Random forest	53	59.6
K-nearest neighbors	53	59.6
Stochastic gradient descent	52	58.4

would be a need to enhance the selection of data variables for scoring the manuscripts with a greater focus towards objective variables. Other limitations included small sample size and possible interrater variability in scoring individual manuscripts.

### References

1. Checco A, Bracciale L, Loret P, Pinfield S, Bianchi G. AI-assisted peer review. *Humanit Soc Sci Commun*. 2021;8(1):25. doi.org/10.1057/s41599-020-00703-8
2. Heaven D. AI peer reviewers unleashed to ease publishing grind. *Nature*. 2018;563:609-610. doi.org/10.1038/d41586-018-07245-9

<sup>1</sup>Department of Surgery, University of Nairobi, Nairobi, Kenya, drenk10@gmail.com; <sup>2</sup>*Annals of African Surgery*, Nairobi, Kenya

**Conflict of Interest Disclosures** Dr Kigera is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

### Authorship and Contributorship

#### Association Between Gift Authorship and Peer-Reviewed Publications and Research Funding Awarded Through Competitive Grants in Different Disciplines

Eric A. Fong,<sup>1</sup> Yeolan Lee,<sup>1</sup> Allen W. Wilhite<sup>2</sup>

**Objective** Recent studies have found gift authorship to be one of the most prevalent types of research misbehavior,<sup>1,2</sup> and prior research reported gifting rates between 20% and 50%.<sup>3</sup> This study suggests one reason for its popularity may be due to its association with publications and grant funding, 2 commonly used measures of academic success.

**Design** Using a convenience sample, surveys were sent to approximately 118,400 faculty from the top 200 universities listed in *US News & World Report*. Respondents self-reported their academic rank and discipline as well as their publications, grant funding, and experience with gift authorship, in the preceding 5 years. They were also asked the main reason why their most recent gift authorship was given to a particular individual (ie, the gift recipient). Surveys were distributed in 2012 to 2014; this study used responses to 11 of the 28 questions, excluding questions that addressed other forms of misconduct. Data were analyzed using 2-stage least squares regression: the first stage explored who was more likely to add a gift author and the second stage measured the

associations between gift authors, publications, and grant funding. Responses were then separated into 2 gift recipient types: administrative roles or nonadministrative colleagues. The associations of gifting with publications and with grant funding were then recalculated and compared.

**Results** The response rate to the survey was 10.4% (12,317/118,400); this was about what was expected given the length of the cover email, the length of the survey, and the potential the emails were directed to junk mail, filtered as spam, etc. Of the 12,317 responses, incomplete surveys and responses from unrelated disciplines reduced the usable sample to 10,689 responses. Of the respondents who responded to rank and sex questions, 33.2% (3956/11,919) were female, 21.2% (1878/8838) were assistant professors, 27.4% (2418/8838) were associate professors, 46.2% (4080/8838) were professors, and the balance consisted of other ranks. Thirty-five percent (3749/10,698) of respondents admitted to participating in gift authorship. The results showed a positive association between the practice of gift authorship and a scholar's number of publications and their research funding (**Table 28**). Furthermore, adding an administrator as a gift author was correlated with larger increases in publications and more grant funding than gifting nonadministrator colleagues.

**Conclusions** The association between participating in gift authorship and academic success in the US, documented in this study, suggested that it could be difficult for the academy to reduce the use of gift authorship unless tangible costs and enforcement of those costs are in place.

## References

- Reisig MD, Holtfreter K, Berzofsky ME. Assessing the perceived prevalence of research fraud among faculty at

**Table 28. Association Between Gift Authorship and Publications and Grant Funding**

	Publications, coefficient (SE) <sup>a</sup>	Grant funding, coefficient (SE) <sup>a</sup>
Total, No.	8295	5625
Constant	1.899 (0.041) <sup>b</sup>	11.983 (0.116) <sup>b</sup>
Gifted authorship	0.503 (0.094) <sup>b</sup>	0.781 (0.246) <sup>b</sup>
Assistant professor	-0.608 (0.028) <sup>b</sup>	-1.451 (0.074) <sup>b</sup>
Associate professor	-0.335 (0.023) <sup>b</sup>	-0.667 (0.059) <sup>b</sup>
Lecturer	-1.208 (0.052) <sup>b</sup>	-2.069 (0.189) <sup>b</sup>
Clinical faculty	-0.667 (0.069) <sup>b</sup>	-1.025 (0.242) <sup>b</sup>
Research faculty	-0.031 (0.053)	0.073 (0.137)
Female	-0.169 (0.022) <sup>b</sup>	-0.179 (0.058) <sup>b</sup>
No. of authors	0.002 (0.000) <sup>b</sup>	0.001 (0.000) <sup>c</sup>
$\chi^2$ Value	2578.5	2071.1

<sup>a</sup>The dependent variable in the first data column was in publications in the last 5 years, and for the second data column was in grant funding received in the last 5 years.

<sup>b</sup>Significance at the .05 level.

<sup>c</sup>Significance at the .05 level. The disciplines (accounting, finance, management, marketing, information systems, political science, psychology, sociology, biology, chemistry, computer science, ecology, engineering, mathematics, physics, and economics as the omitted reference group) were included in the regression.

research-intensive universities in the USA. *Account Res.* 2020;27(7):457-475. doi:10.1080/08989621.2020.1772060

2. Bouter LM, Tijdink J, Axelsen N, Matinson BC, Ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conference on Research Integrity. *Res Integ Peer Rev.* 2016;1:17. doi:10.1186/s41073-016-0024-5

3. Elliott KC, Settles IH, Montgomery GM, Brassel ST, Cheruvellil KS, Soranno PA. Honorary authorship practices in environmental science teams: structural and cultural factors and solutions. *Account Res.* 2017; 24(2):80-98. doi:10.1080/08989621.2016.1251320

<sup>1</sup>Department of Management, Marketing, and Information Systems, University of Alabama in Huntsville, AL, USA, eric.fong@uah.edu;

<sup>2</sup>Department of Accounting, Economics, and Finance, University of Alabama in Huntsville, AL, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This study was supported by the Office of Research Integrity through the Department of Health and Human Services (grant ORIIR130003).

**Additional Information** The content is that of the authors and does not necessarily represent the official views of, nor an endorsement by, the Health Resources and Services Administration, Department of Health and Human Services, or the US government.

## Numbers and Trends in Authorship of Published Case Reports in Plastic Surgery Journals, 1956-2018

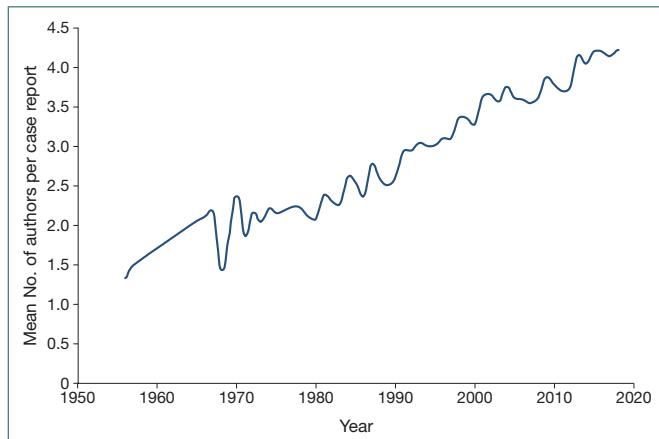
Marios Papadakis<sup>1</sup>

**Background** Although some studies reporting authorship proliferation in plastic surgery exist,<sup>1</sup> there are no studies about case reports. Moreover, most studies report only 1 or 2 journals. The aim of this study was to investigate the evolution of authorship in MEDLINE-indexed case reports published in the majority of plastic surgery journals over time.

**Design** A list of the top plastic surgery journals according to the Scimago Journal & Country Rank<sup>2</sup> was made. Excluded were journals that do not consider case reports, journals that limit the number of authors, and wound journals owing to the broad spectrum of specialties publishing in these journals. Case reports published in journals that later stopped considering case reports were included. Each journal was separately searched and filtered in PubMed for case reports published between 1956 and 2018.

**Results** A total of 13,960 case reports published in 14 plastic surgery journals between 1956 and 2018 were analyzed. The number of case reports published reached its peak in 2009, and this number has been decreasing since 2010. The overall percentage of case reports among all manuscripts published was 22%, ranging from 3% (in *JAMA Facial Plastic Surgery*) to 48% (in *Ophthalmic Plastic and Reconstructive Surgery*). The mean (SD) number of authors was 3.4 (1.7). The mean

**Figure 16. Case Report Authorship, 1956 to 2018**



number of authors per manuscript increased from 1.3 in 1956 to 2.6 in 1990 and 4.2 in 2018 (**Figure 16**). The most common number of authors was 3 (found in 25% of case reports), followed by 2 (21%). Overall, 87% of the case reports had between 1 and 5 authors, whereas only 0.2% had 11 or more authors. One case report had 20 authors and another one had 19 authors, whereas no case report with 16 to 18 authors was found.

**Conclusions** Plastic surgery remains a field in which case reports are published frequently (22% overall), although there has been a decreasing trend during the last 10 years. The trend of authorship proliferation reported in original articles submitted in plastic surgery journals is also observed in case reports, with a current mean number of 4.2 authors per article. However, the increase is not as extreme as in other fields.<sup>3</sup>

## References

1. Durani P, Rimouche S, Ross G. "How many plastic surgeons does it take to write a research article?" authorship proliferation in and internationalisation of the plastic surgery literature. *J Plast Reconstr Aesthet Surg.* 2007;60(8):956-957. doi:10.1016/j.bjps.2006.08.002
2. Scimago Lab. Scimago Journal & Country Rank. Accessed July 11, 2022. <https://www.scimagojr.com/>
3. Papadakis M. How many scientists does it take to write a COVID-19 case report? *Account Res.* 2021;28(3):186-190. doi:10.1080/08989621.2020.1821369

<sup>1</sup>Department of Surgery II, University Witten-Herdecke, Witten, Germany, marios\_papadakis@yahoo.gr

**Conflict of Interest Disclosures** None reported.

## A Systematic Review of Survey Research of Honorary Authorship in Health Sciences

Reint A. Meursinge Reynders,<sup>1,2</sup> Gerben ter Riet,<sup>3,4</sup> Nicola Di Girolamo,<sup>5,6</sup> Davide Cavagnetto,<sup>1,2</sup> Mario Malički<sup>7</sup>

**Objective** Perceived honorary authorship refers to the perception or opinion of survey respondents that 1 or more co-authors should not have been included as author(s) of a

publication because they made no or insufficient contributions to qualify as authors. In this systematic review, the prevalence of perceived honorary authorship and 4 other honorary authorship issues in health sciences were assessed. These issues represent a subgroup of the research questions reported in a previously published protocol<sup>1</sup> and build on prior research on honorary authorship.<sup>2,3</sup>

**Design** Surveys of authors of scientific publications on 5 honorary authorship issues in health sciences in any language were eligible. A search from inception until December 8, 2021, was done in PubMed, Lens.org, and Dimensions.ai. Two authors conducted the study selection and data extraction procedures independently. The prevalence of researchers perceiving other coauthor(s) as honorary author(s) on a publication (review item 1) and the prevalence of researchers having been approached by others to include honorary author(s) on a publication (review item 2) were the primary outcomes. Outcomes were exclusively based on what was asked in the questionnaires. The methodological quality of surveys was assessed with a self-developed 14-item checklist (available in the protocol).<sup>1</sup> Qualitative and quantitative syntheses were conducted and Metaprop in Stata was used to perform the meta-analyses (random effects model) of the prevalence statistics. Double arcsine transformation was performed prior to statistical pooling.

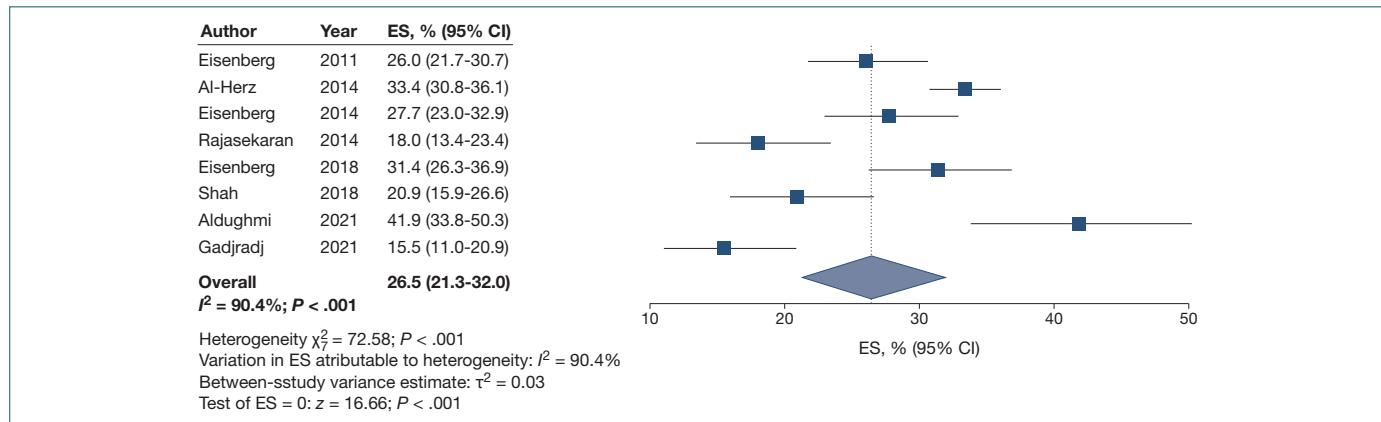
**Results** After removal of duplicates, 1220 articles were screened, of which 8 surveys were eligible for review item 1. No surveys were eligible for review item 2 nor for any of the other 3 issues sought on perceived honorary authorship. Many surveys were excluded because of spin, ie, definitions of perceived honorary authorship in the main text were eligible, but not those used in the questionnaires. The pooled response rate on perceived honorary authorship in the 8 eligible surveys (15,553 contacted authors) was 24.8% (95% CI, 19.9%-29.9%). The pooled prevalence of perceived honorary authorship on 3132 survey respondents was 26.5% (95% CI, 21.3%-32.0%) (**Figure 17**). The low *P* value and large  $\chi^2$  ( $\chi^2_7 = 72.58$ ;  $P < .001$ ) provide evidence of heterogeneity, and the high  $I^2$  ( $I^2 = 90.4\%$ ) indicates considerable inconsistency across the prevalence statistics of the surveys. Characteristics of nonresponders were not reported in any of the eligible surveys. The methodological quality was critically low in all eligible surveys.

**Conclusions** This systematic review found that 26.5% of health scientists perceived having at least 1 honorary author in at least 1 of their publications. However, this estimate should be interpreted with caution because of high risk of bias, considerable heterogeneity, and numerous uncertainties. Future studies should focus on interventions to prevent honorary authorship.<sup>3</sup>

## References

1. Meursinge Reynders R, Ter Riet G, Di Girolamo N, Malički M. Honorary authorship in health sciences: a protocol for a systematic review of survey research. *Syst Rev.* 2022;11(1):57. doi:10.1186/s13643-022-01928-1

**Figure 17. Meta-analysis of Surveys on the Prevalence of Perceived Honorary Authorship**



ES indicates effect size.

2. Marušić A, Bošnjak L, Jerončić A. A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS One.* 2011;6(9):e23477. doi:10.1371/journal.pone.0023477

3. Patience GS, Galli F, Patience PA, Boffito DC. Intellectual contributions meriting authorship: survey results from the top cited authors across all science categories. *PLoS One.* 2019;14(1):e0198117. doi:10.1371/journal.pone.0198117

<sup>1</sup>Department of Oral and Maxillofacial Surgery, Amsterdam University Medical Center, Amsterdam, the Netherlands, reyndersmail@gmail.com; <sup>2</sup>Studio di Ortodonzia, Milan, Italy; <sup>3</sup>Urban Vitality Centre of Expertise, Amsterdam University of Applied Sciences, Amsterdam, the Netherlands; <sup>4</sup>Department of Cardiology, Amsterdam University Medical Center, Amsterdam, the Netherlands; <sup>5</sup>Center for Veterinary Health Sciences, Oklahoma State University, Stillwater, OK, USA; <sup>6</sup>EBMVet, Cremona, Italy; <sup>7</sup>Meta-research Innovation Center at Stanford, Stanford University, Stanford, CA, USA

**Conflict of Interest Disclosures** None reported.

## Conflict of Interest

### Accuracy of Conflict of Interest Disclosure Among Australian Clinical Trial Authors

Lorelie Flood,<sup>1</sup> Barbara Mintzes,<sup>1</sup> Kellia Chiu,<sup>1</sup> Zhaoli Dai,<sup>1,2</sup> Emily A. Karanges,<sup>3,4</sup> Bennett Holman<sup>5,6</sup>

**Objective** Authors' financial ties with pharmaceutical companies can affect the design, conduct, and reporting of clinical trials of drug treatments.<sup>1</sup> Although disclosure does not eliminate conflicts of interest (COI), it allows readers and reviewers to consider potential effects. There has been little research on the accuracy of authors' COI disclosures outside the US.<sup>2,3</sup> However, since 2015, Medicines Australia (MA), Australia's research-based pharmaceutical industry association, has required member companies to report payments provided to individual clinicians. This study assessed how often Australian clinical trial authors accurately report pharmaceutical industry financing by comparing authors' self-reported COI in published articles with MA payments data, using the International Committee of Medical

Journal Editors (ICMJE) criteria to assess adequacy of disclosure. A secondary analysis compared numbers of authors per trial with inaccurate disclosures in journals with author instructions consistent with ICMJE standards, as compared with journals with weaker disclosure instructions.

**Design** This was a cross-sectional study. To identify Australian authors of recent randomized clinical trials (RCTs), Ovid Medline was searched from January to August 2020 using a Cochrane RCT search filter, limited by mention of *Australia* anywhere in the text, including author affiliations. RCTs that tested prescription-only medicines and vaccines in clinical populations were included. Two researchers independently compared authors' disclosures in included trials with information in the MA database, assessing payments from companies operating in the relevant commercial space within a 3-year period before article submission. Relevant commercial space was defined as marketing products for the same condition or therapeutic class, as per ICMJE criteria. Journals' instructions for authors on COI disclosure were classified according to consistency with ICMJE criteria. To compare Australian and US authors' reporting rates, the subset of trials with US authors was identified, and US Open Payments data on general payments (excluding food and beverage) were used to match MA data.

**Results** Of 583 unique identified records, 120 met inclusion criteria as drug trials with 1 or more Australian authors. In total, 56 of 120 trials (47%) had 1 or more authors with undisclosed COI, and 78 of 323 Australian authors (24%) had undisclosed COI (89 nondisclosures). Among the remaining authors, 129 of 323 (40%) had accurate declarations. We could not assess disclosure accuracy for 116 of 323 nonclinicians (36%). The most common type of nondisclosure was incorrectly declaring no COI (46 [52%]), followed by partial disclosures (39 [44%]). The median value of undisclosed payments was \$8,944 (range, \$140-\$97,600) Australian dollars. Author nondisclosure rate per trial was similar whether or not journals applied ICMJE criteria: 40% vs 45% ( $P = .51$ ).

**Conclusions** In this sample of recent RCTs with Australian authors, inaccurate and incomplete COI declarations were

common. These discrepancies highlight the need for more transparent and comprehensive COI reporting.

## References

1. Fabbri A, Lai A, Grundy Q, Bero LA. The influence of industry sponsorship on the research agenda: a scoping review. *Am J Public Health*. 2018;108:e9-e16. Medline: 30252531 doi:10.2105/AJPH.2018.304677
2. Taheri C, Kirubarajan A, Li X, et al. Discrepancies in self-reported financial conflicts of interest disclosures by physicians: a systematic review. *BMJ Open*. 2021;11:e045306. doi:10.1136/bmjopen-2020-045306
3. Rasmussen K, Schroll J, Gøtzsche PC, Lundh A. Underreporting of conflicts of interest among trialists: a cross-sectional study. *J R Soc Med*. 2015;108(3):101-107. doi:10.1177/0141076814557878

<sup>1</sup>Charles Perkins Centre, School of Pharmacy, Faculty of Medicine and Health, The University of Sydney, Camperdown, New South Wales, Australia, Barbara.mintzes@sydney.edu.au; <sup>2</sup>College of Medicine and Public Health, Flinders University, Bedford Park, South Australia, Australia; <sup>3</sup>Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia; <sup>4</sup>Oxygen, Melbourne, Victoria, Australia; <sup>5</sup>Underwood International College, Yonsei University, Incheon, Korea; <sup>6</sup>Faculty of Humanities, University of Johannesburg, Johannesburg, South Africa

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was funded by grant AKF2020192 from the Australia Korea Foundation, Department of Foreign Affairs and Trade, and by a University of Sydney and Yonsei University Partnership Collaboration Award.

**Role of the Funders/Sponsor** Neither funder had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Additional Information** Bennett Holman is a co-corresponding author.

## Conflicts of Interest and the Role of Funders and Authors in Clinical Trials Included in Cochrane Reviews

Erlend Faltinsen,<sup>1,2</sup> Adnan Todorovac,<sup>3</sup> Isabelle Boutron,<sup>4</sup> Lesley Stewart,<sup>5</sup> Asbjørn Hróbjartsson,<sup>1,2</sup> Andreas Lundh<sup>1,2,6</sup>

**Objective** In Cochrane reviews, trial authors' conflicts of interest are often not reported and trial funding information is sometimes missing.<sup>1,2</sup> This study assessed (1) the proportion of Cochrane reviews reporting trial funding and authors' conflicts of interest and (2) whether accessing the main trial publication and searching other information sources could identify additional information on funding, conflicts of interest, and the role of funders and authors.

**Design** In a cross-sectional study, 1 index trial was randomly included from the primary meta-analyses of 100 Cochrane reviews (October to December 2020). Two authors independently extracted trial characteristics, funding and conflict of interest information, and the role of funders and

authors from the reviews and main trial publications, including conflict of interest disclosure forms. Other sources (eg, trial protocols and registry information) were also searched to retrieve additional information, and the time this took was noted. The proportion of Cochrane reviews and main trial publications reporting trial funding and conflict of interest information and role of funders and authors were calculated, as was the proportion of trials in which additional information was found by searching other sources.

**Results** The included trials were published from 1975 to 2020 (median: 2011), and 47 were drug trials. Sixty-eight reviews reported trial funding, and 25 reported trial authors' conflicts of interest. Accessing the main trial publication led to identification of funding in 16 additional trials and conflict of interest information in 38 additional trials. In trials in which funders or trial authors had conflicts of interest, their roles were sufficiently reported in 20 of 36 (56%) and 20 of 30 (67%) main trial publications, respectively. It took approximately 9 minutes (range, 2-28 minutes) per trial publication to extract information. When searching other sources, additional information on funding was found in 2 trials and authors' conflicts of interest in 13 trials, and it took approximately 22 minutes (range, 4-97 minutes) per trial to extract information. Trial registries and other publications by trial authors were the information sources that most frequently contained additional information.

**Conclusions** One-third of recent Cochrane reviews did not report funding of a randomly selected included trial, and three-quarters did not report trial authors' conflicts of interest despite the information often being reported in the main trial publication. Review authors should systematically access and read main trial publications and disclosure statements and consider searching other information sources.

## References

1. Turner K, Carboni-Jiménez, Benea C, et al. Reporting of drug trial funding sources and author financial conflicts of interest in Cochrane and non-Cochrane meta-analyses: a cross-sectional study. *BMJ Open*. 2020;10:e035633. doi:10.1136/bmjopen-2019-035633
2. Roseman M, Turner EH, Lexchin J, et al. Reporting of conflicts of interest from drug trials in Cochrane reviews: cross sectional study. *BMJ*. 2012;345:e5155. doi:10.1136/bmj.e5155

<sup>1</sup>Centre for Evidence-Based Medicine Odense and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark, egfaltinsen@health.sdu.dk; <sup>2</sup>Open Patient Data Explorative Network, Odense University Hospital, Odense, Denmark; <sup>3</sup>Psychiatric Hospital Esbjerg, Region of Southern Denmark, Esbjerg, Denmark; <sup>4</sup>Université de Paris, CRESS, Inserm, INRA, Paris, France; <sup>5</sup>Centre for Reviews and Dissemination, University of York, York, UK; <sup>6</sup>Department of Infectious Diseases, Hvidovre Hospital, Hvidovre, Denmark

**Conflict of Interest Disclosures** Isabelle Boutron, Lesley Stewart, Asbjørn Hróbjartsson, and Andreas Lundh are members of the TACIT Steering Group involved in developing a Tool for Addressing Conflicts of Interest in Trials. Isabelle Boutron is a

**Table 29. Financial Conflict of Interest by 6 Common Diagnoses in Primary Care and Year of 746 Systematic Review Publications**

fCOI category	Systematic review and diagnoses, No. (%)													
	COPD		T2D		HT		Dementia		MDD		OA		Total	
	2010 (n = 28)	2019 (n = 61)	2010 (n = 25)	2019 (n = 161)	2010 (n = 29)	2019 (n = 77)	2010 (n = 16)	2019 (n = 83)	2010 (n = 35)	2019 (n = 84)	2010 (n = 23)	2019 (n = 124)	2010 (n = 156)	2019 (n = 590)
Yes	8 (29)	13 (21)	5 (20)	28 (17)	5 (17)	5 (6)	5 (31)	8 (10)	12 (34)	27 (32)	3 (13)	28 (23)	38 (24)	109 (18)
No	12 (43)	40 (66)	6 (24)	107 (66)	17 (59)	44 (57)	3 (19)	60 (72)	15 (43)	43 (51)	8 (35)	68 (55)	61 (39)	362 (61)
Information missing	8 (29)	8 (13)	14 (56)	26 (17)	7 (24)	28 (36)	8 (50)	15 (18)	8 (23)	14 (17)	12 (52)	28 (23)	57 (37)	119 (20)

Abbreviations: COPD, chronic obstructive pulmonary disease; fCOI, financial conflict of interest; HT, hypertension; MDD, major depressive disorder; OA, osteoarthritis; T2D, type 2 diabetes.

member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. No other disclosures were reported.

**Additional Information** Andreas Lundh is a co-corresponding author.

### Conflict of Interest in Published Systematic Reviews on Interventions for 6 Common Clinical Diagnoses, 2010–2019

Marek Czajkowski,<sup>1,2</sup> Alexandra Snellman,<sup>1,2</sup> Louise Olsson<sup>1,2</sup>

**Objective** Conflicts of interest in systematic reviews (SRs) have gained recognition,<sup>1,2</sup> but overall knowledge is limited, in particular from a clinical perspective. This study addresses financial conflict of interest (fCOI) statements in SRs relevant to primary care.

**Design** A cross-sectional study of SRs published in 2010 and 2019 that evaluated interventions for 6 common diagnoses was designed. The diagnoses covered a substantial range of primary care specialties: chronic obstructive pulmonary disease, type 2 diabetes, hypertension, dementia, major depressive disorder, and osteoarthritis. PubMed was searched by a librarian in June 2021. Two independent researchers selected relevant SRs; any disagreements were resolved in consensus. Basic characteristics and data on fCOI and funding were extracted from full-text versions of SRs. The SRs of authors disclosing fCOI or reporting funding from the industry were categorized as “yes.” The SRs of authors declaring no fCOI and reporting no funding from the industry were categorized as “no.” Missing information either on authors’ disclosure of conflicting interests or information on funding of the SRs was categorized as “information missing.” A random sample of 50 SRs per fCOI category (“yes,” “no,” or “information missing”) was drawn, and data on fCOI and funding of the primary studies reported in these 150 SRs were extracted.

**Results** Of 2234 initially retrieved records, 746 (33%) relevant SRs were included. Diabetes and pharmacological interventions were the most common topics: 186 of 746 SRs (25%) for diabetes and 247 of 746 (33%) SRs for pharmacological interventions. Most often the first author was affiliated with an institution in China (130 [17%]). The number of relevant SRs increased 4-fold, from 156 in 2010 to 590 in 2019. In 2010, 38 (24%) SRs fulfilled the criteria for the fCOI “yes” category vs 109 (18%) in 2019 (**Table 29**). For

osteoarthritis, the pattern was the opposite, with 3 (13%) categorized as “yes” in 2010 and 28 (23%) in 2019. The largest proportion of the fCOI “yes” category was noticed for depression, with 12 (34%) SRs in 2010 and 27 (32%) in 2019. In 2010, 57 (37%) SRs were categorized as “information missing” vs 120 (20%) in 2019. The only exception was for hypertension, with 7 (24%) SRs in 2010 vs 28 (36%) in 2019. Of the random subsample of 150 SRs, 30 (20%) reported data on fCOI or funding for the included primary studies.

**Conclusions** The proportion of SRs disclosing fCOI or funding from industry was lower in 2019 compared with 2010 but was still missing in every fifth SR, with a direct clinical association with primary care. Risk factors for these shortcomings need to be identified, and given this, policies of journals are notable. Only a few SRs reported data on fCOI among the included primary studies.

### References

1. Hakoum MB, Anouti S, Al-Gibbawi M, et al. Reporting of financial and non-financial conflicts of interest by authors of systematic reviews: a methodological survey. *BMJ Open*. 2016;6:e011997. doi:10.1136/bmjopen-2016-011997
2. Hansen C, Lundh A, Rasmussen K, Hróbjartsson A. Financial conflicts of interest in systematic reviews: associations with results, conclusions, and methodological quality. *Cochrane Database Syst Rev*. 2019;8(8):MR000047. doi:10.1002/14651858.MR000047.pub2

<sup>1</sup>School of Medical Sciences, Örebro University, Örebro, Sweden, marek.czajkowski@oru.se; <sup>2</sup>Centre for Assessment of Medical Technology in Örebro, Örebro University Hospital, Örebro, Sweden

**Conflict of Interest Disclosures** None reported.

**Funding/Support** The project was supported financially by the Center for Clinical Research Dalarna, Falun, Sweden.

**Role of the Funder/Sponsor** The funder did not have any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Data Presentation and Graphical Display

### Redesigning Web-based Presentation of Agency for Healthcare Research and Quality Evidence-based Practice Center Program Systematic Reviews

Celia V. Fiordalisi,<sup>1</sup> Edwin Reid,<sup>1</sup> Haley K. Holmer,<sup>1</sup> Edi E. Kuhn<sup>1</sup>

**Objective** The Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) Program produces rigorous and comprehensive systematic reviews that are posted on the Effective Health Care (EHC) website. To improve presentation, user experience, and accessibility of reports and their findings, a redesign of the EHC website was implemented to display important systematic review elements through interactive tools presented on a clean and easy-to-use webpage.

**Design** The interactive site was initially designed by the Scientific Resource Center (SRC) before being piloted by the American Institutes for Research among clinical and other stakeholders within 6 large learning health systems. The SRC incorporated findings from pilot activities to update the site design, and the EPC Program is now implementing this new web-based design that prioritizes interactive data presentation and end-user inquiry-based exploration of findings.

**Results** To date, 7 EPC systematic reviews had been transformed into interactive formats and were live on the EHC website, with additional reports being added as available. Three elements improved accessibility of report findings and are key features in implementation of the site redesign: Main Findings (summary bullet points), the Visual Dashboard (interactive graphs and tables designed in Tableau), and the Report Snapshot. The Main Findings allow end users to get the bottom line with minimal time investment. The Visual Dashboard offers customizable data visualizations, allowing end users to explore data relevant to their clinical decision-making or research questions. For users who prefer a more text-based presentation, the Report Snapshot displays data and summary statements within a table. Both the Visual Dashboard and the Report Snapshot hyperlink specific data points to the underlying studies on PubMed. These design elements address end users' need to quickly identify key messages and evaluate the data according to their own needs and questions, whether these users are health system leaders, practicing clinicians, guideline developers, or researchers.

**Conclusions** The EHC website redesign offers users multiple methods to explore the findings from AHRQ EPC reports in an accessible and dynamic format. The implementation of the redesign will incorporate the peer review process to improve the quality of reporting. Evidence producers of all types should begin to better meet the informational needs of evidence consumers by moving away

from flat-file presentation of findings toward more interactive and web-based displays.

<sup>1</sup>Portland VA Research Foundation, Portland, OR, USA, haley.holmer@va.gov

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was funded by the AHRQ EHC Program through a contract to the SRC (IHSA 2902017003C). The authors of this manuscript are responsible for its content. Statements in the manuscript do not necessarily represent the official views of or imply endorsement by the AHRQ.

## Data Sharing and Access

### Perspectives on Responsibilities in Receipt and Secondary Use of Data in Health Research

Kylie E. Hunter,<sup>1</sup> Aidan C. Tan,<sup>1</sup> Angela C. Webster,<sup>1</sup> Daniel G. Hamilton,<sup>2,3</sup> Myra Cheng,<sup>4</sup> Lee Jones,<sup>5</sup> Sol Libesman,<sup>1</sup> Salma Fahridin,<sup>1</sup> Antonio Laguna Camacho,<sup>6</sup> Rui Wang,<sup>7</sup> Anna Lene Seidler<sup>1</sup>

**Objective** Despite strong in-principle support for the concept of data sharing for health research, in practice it is often difficult to find, access, and reuse data.<sup>1</sup> In the absence of agreed-on global standards, this study sought to determine the responsibilities of recipients of health research data and propose how these responsibilities can be met.

**Design** A qualitative study involving an online focus group was conducted in December 2021 at the Association for Interdisciplinary Meta-research and Open Science (AIMOS) conference. All conference delegates were eligible to attend, and targeted invitations were sent to known data-sharing experts. The conference was open to all and offered free attendance. The focus group involved discussion of 3 case studies of different data-sharing scenarios in health research (an individual participant data meta-analysis, study replication, and secondary analyses) and a general discussion prompted by key questions. Participants contributed by speaking in the video call or by typing in real time on shared Google documents. Afterward, notes and recording transcripts were collated into categories using thematic analysis and shared with attendees for review and further input. Primary outcomes were the responsibilities of data recipients across the design, conduct, analyses, and reporting stages of their research. Secondary outcomes included how data providers may support data recipients to meet these responsibilities.

**Results** The 2-hour focus group discussion was attended by 16 conference delegates (including 3 facilitators; 11 delegates agreed to coauthor this abstract). Although AIMOS is a multidisciplinary conference, most attendees had health care-related roles across various fields, including epidemiology, statistics, evidence synthesis, policy, ethics, public health, data management, oncology, psychology, nutrition, clinical trials, and administration. Most participants cited a university as their primary employer.

**Table 30. Responsibilities of Data Recipients to Data Providers Across Study Stages**

Study stage	Responsibilities of data recipients
Design	Seek ethical approval or waiver of consent for use of data and share with data provider
	Codraft data sharing or access agreement and terms with data provider
	Offer data provider opportunity to collaborate and give input on protocol and analysis plan
	Check legislative or other regulatory requirements of the data provider's state or country
	Openly publish or prospectively register research protocol and share with data provider
	Prepare and share data management plan with data provider
	Allow data provider to publish first results before sharing (or explicitly agree otherwise)
Conduct	Request deidentified data (and dictionary) in least effort format; discern reidentification risk
	Clarify data elements and assure understanding of data
	Ensure sufficient resources to process, check, clean, and verify data in a reasonable timeframe
	Continue to adhere to local codes of conduct
Analysis	Demonstrate requisite clinical knowledge, data management, and analytical skills
	Share the analytic code in a public repository
	Only include in analyses data and studies that meet prespecified quality and integrity standards
Report	Offer opportunity for data provider to review results and manuscript prior to publication
	Offer authorship, acknowledgment, or credit on study outputs when appropriate

Analyses revealed several recurring themes across the data-sharing scenarios, which were grouped into recommendations (**Table 30**). Recommendations included that data recipients need to prioritize the protection of participant privacy and should proactively share a secure data management plan and evidence of ethical approval with the data provider. Additionally, data recipients should allay concerns about potential data misuse by demonstrating they have sufficient resources and expertise to process, check, and analyze data. They should also protect the interests of data providers by allowing them to publish their results before sharing data, offering streamlined data-sharing pathways, and inviting them to contribute to related outputs. Data providers could support recipients by planning for data sharing during study design, including funding support, local legislation, intellectual property, commercial-in-confidence information, and by preparing a data dictionary.

**Conclusions** This study provides clarity around responsibilities of data recipients to address common concerns of data providers on data misuse and privacy. Several recommendations were derived on how data providers can support recipients to bridge the gap between high support for data sharing and low in-practice compliance.

## Reference

1. Tan AC, Askie LM, Hunter KE, Barba A, Simes RJ, Seidler AL. Data sharing—trialists' plans at registration, attitudes, barriers and facilitators: a cohort study and cross-sectional survey. *Res Syn Meth.* 2021; 1-17. doi:10.1002/jrsm.1500

<sup>1</sup>National Health and Medical Research Council (NHMRC) Clinical Trials Centre, University of Sydney, Camperdown, NSW, Australia, kylie.hunter@sydney.edu.au; <sup>2</sup>MetaMelb Research Group, School of BioSciences, The University of Melbourne, Melbourne, Australia;

<sup>3</sup>Melbourne Medical School, Faculty of Medicine, Dentistry & Health Sciences, The University of Melbourne, Melbourne, Australia;

<sup>4</sup>Australian Research Data Commons , Australia; <sup>5</sup>School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia; <sup>6</sup>Autonomous University of the State of Mexico, Toluca, Mexico; <sup>7</sup>Department of Obstetrics & Gynaecology, Monash Health, Monash University, Melbourne, Australia

**Conflict of Interest Disclosures** Kylie E. Hunter and Antonio Laguna Camacho lead or colead several large individual participant data meta-analyses. Kylie E. Hunter and Angela C. Webster are associate convenors of the Cochrane Prospective Meta-Analysis Methods Group, and Antonio Laguna Camacho is coconvenor. Daniel G. Hamilton is a member of the Association for Interdisciplinary Meta-research and Open Science board. No other disclosures were reported.

**Funding/Support** Kylie E. Hunter received research funding support via 2 scholarships administered by the University of Sydney (Postgraduate Research Supplementary Scholarship in Methods Development [SC3504] and Research Training Program Stipend [SC3227]). Anna Lene Seidler received a National Health and Medical Research Council Investigator Grant (GNT2009432).

**Role of the Funder/Sponsor** The funders had no role in any of the following: design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgment** We warmly acknowledge the contributions of all those who attended the focus group discussion.

## Dissemination of Information

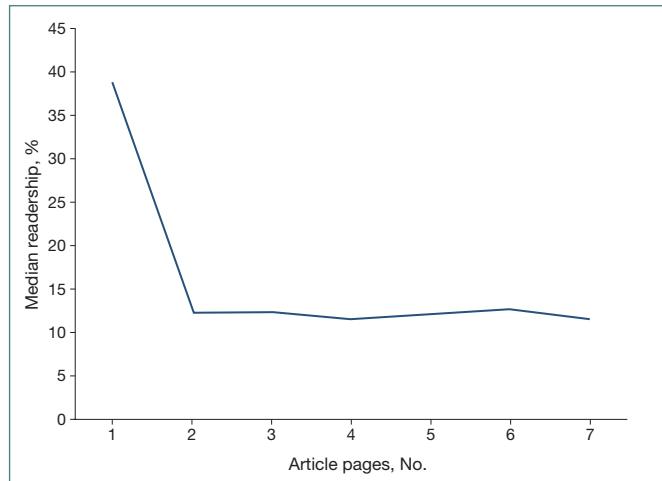
### Primary Care Physician Readership Practices of the Printed Versions of *Deutsches Ärzteblatt*

Christopher Baethge,<sup>1,2</sup> Jeremy Franklin<sup>3</sup>

**Objective** How medical journal audiences read scientific articles is of key interest to authors and editors, but there are few publications on reading patterns. To better understand reading behavior, surveys of readers of a print general medical journal conducted with physicians in private practice were analyzed, with a focus on reading rates, popularity of different kinds of articles, attraction of tables or figures, and how readership declined with increasing numbers of pages in an article.

**Design** This was a survey study of readers of *Deutsches Ärzteblatt*, the journal of the German Medical Association and the National Association of Statutory Health Insurance Physicians. During regular surveys between 2000 and 2019, changing groups of general practitioners and internists in private practice marked pages they had read in a print journal

**Figure 18. Readership of Full-Length Original Articles and Review Articles in a General Medical Journal**



issue. Here, we present results from 57 survey waves. Using multivariable logistic regression, associations of formal page characteristics (independent variables: right vs left page, first vs following page within an article, presence of tables or text boxes, publication year) with reading status of a page (dependent variable: yes/no) were analyzed.

**Results** The median number of participants per survey wave was 49 (IQR, 35-77), the mean age across groups was 54 years, and a total of 1104 pages were surveyed. In all waves combined, the median readership per page was 14.1% (IQR, 6.7%-25.0%), with no trend across the 20 years covered. For full-length articles, the mean (SD) readership peaked at page 1 at 38.7% (11.5%), decreased to 12.3% (8.7%) at page 2, and plateaued at approximately 12% for pages 3 to 7 (**Figure 18**). Reference lists were read less frequently (median, 0% [IQR, 0%-1.6%]). Pages with tables were marked more often than those without (median, 12.5% [IQR, 6.3%-20.1%]), less so for pages with figures (median, 9.4% [IQR, 4.5%-17.6%]). Readers more often read right than left pages, but after removing first pages from the analysis, the difference narrowed (median, 11.3% [IQR, 5.6%-21.2%] vs median, 10.1% [IQR, 5.0%-17.4%]). On multivariable analysis, reading status was independently associated with first pages (OR, 4.70 [95% CI, 4.10-5.38]), right pages (OR, 1.14 [95% CI, 1.04-1.25]), and presence of tables or text boxes (OR, 1.32 [95% CI, 1.16-1.49]). Physicians more often read review articles (median, 17.0% [IQR, 12.3%-24.0%]) than original articles (median, 12.5% [IQR, 10.0%-15.8%]) and letters (median, 11.7% [IQR, 7.8-16.5%]) and read editorials most frequently (median, 29.5% [IQR, 18.8%-34.0%]). By topic, the median (IQR) readership was 18.6% (7.8%-26.9%) for nonsurgery topics (eg, internal medicine, neurology, psychiatry), 12.7% (6.0%-24.1%) for surgery topics (eg, traumatology), and 8.1% (3.1%-15.4%) for cross-disciplinary topics (eg, radiology, microbiology).

**Conclusions** In this study, readership of medical journal articles dropped by two-thirds after page 1 but remained relatively stable thereafter, and approximately 1 in 8 readers

read an entire full-length article. These results apply to primary care physicians and are not representative of the entire audience of a general medical journal.

<sup>1</sup>Deutsches Ärzteblatt and Deutsches Ärzteblatt International, Cologne, Germany, baethge@aerzteblatt.de; <sup>2</sup>Department of Psychiatry and Psychotherapy, Cologne University Medical School, Cologne, Germany; <sup>3</sup>Institute of Medical Statistics, Informatics and Epidemiology, Cologne University Medical School, Cologne, Germany

**Conflict of Interest Disclosures** Christopher Baethge is employed by *Deutsches Ärzteblatt* and is responsible for its scientific content.

**Funding/Support** The surveys analyzed were paid for by Deutscher Ärzteverlag, the publisher of *Deutsches Ärzteblatt*.

## Diversity, Equity, and Inclusion

### Evaluation of Women Representation in National Institutes of Health Study Sections

Cole Wayant,<sup>1</sup> Matt Vassar<sup>2</sup>

**Objective** To study the representation of women scientists and physicians in National Institutes of Health (NIH) study sections.

**Design** In this cross-sectional study, a stratified random sample composed of 15 study sections representing 15 separate NIH institutes and centers was extracted. Rosters from each study section meeting in the years 2011, 2016, and 2021 were downloaded from the NIH Scientific Review Group Roster Index. For each NIH reviewer, professional websites were searched to identify gender identity descriptions. If these descriptions could not be found, genderize.io (Demografix ApS) was used to determine the probability of a name belonging to a man or woman. Names with probability values less than .60 were excluded. The following data points were extracted from each roster: (1) highest academic degree, (2) meeting date, (3) academic rank, (4) type of member (permanent, temporary, chair, scientific review officer, other), (5) institution, and (6) geographic region.<sup>1</sup> Stata version 16.1 was used to conduct  $\chi^2$  tests for 2-group comparisons and a multivariate, binary logistic regression to examine the association of the key factors with the likelihood of being a woman NIH reviewer.

**Results** A total of 3478 total reviewers from 15 study sections in 2011, 2016, and 2021 were identified. Overall, there were 1508 women reviewers (44.6%) and 1970 men reviewers (55.4%). After removing duplicates, there were 1901 unique reviewers, of which 802 (42.2%) were women and 1099 (57.8%) were men. There was no significant difference in the proportion of total and unique women reviewers ( $P = .41$ ). The proportion of women nominally increased each year from 2011 (476 of 1165 [40.9%]) to 2016 (472 of 1123 [42.0%]) to 2021 (560 of 1109 [47.1%]). Multiple regression indicated that women reviewers were less likely to hold medical science degrees. Women were more likely to be from the Southeast and Northeast regions of the US and hold the position of

**Table 31. Binary Logistic Regression Investigating the Key Factors Associated With Woman Representation in National Institutes of Health Study Sections**

Covariate	Women, No./total No. (%)	Odds ratio (95% CI)
<b>Highest degree held</b>		
Doctorate	1116/2333 (48.8)	1 [Reference]
Doctor of medical science	209/635 (32.9)	0.55 (0.46-0.67)
Master's	28/38 (73.7)	3.47 (1.67-7.22)
Combined doctorate–doctor of medicine	135/438 (30.8)	0.51 (0.40-0.62)
Bachelor's	1/1 (100)	1.00
<b>Region of US (or world)</b>		
West	247/656 (37.7)	1 [Reference]
Southwest	142/322 (44.1)	1.27 (0.96-1.67)
Midwest	293/754 (38.9)	1.08 (0.87-1.35)
Southeast	427/870 (49.1)	1.41 (1.13-1.75)
Northeast	368/830 (44.3)	1.39 (1.12-1.72)
Canada/UK	12/13 (92.3)	22.22 (2.85-173.43)
<b>Study section position</b>		
Permanent member	795/1905 (41.7)	1 [Reference]
Chair	86/168 (51.2)	1.50 (1.09-2.07)
Scientific review officer	95/143 (66.4)	2.11 (1.44-3.08)
Temporary member	514/1229 (41.8)	0.97 (0.83-1.12)

Chair or Scientific Review Officer in the study section (**Table 31**).

**Conclusions** This study examining women representation in NIH study sections in 2011, 2016, and 2021 indicates that the proportion of women reviewers increased over time but that several key improvements could be made to further increase representation. In particular, women physicians and physician-scientists were underrepresented in our study. Bias toward women in medicine has been extensively described previously,<sup>2,3</sup> and efforts to identify or overcome limitations to women physician recruitment to NIH study sections should be undertaken. These findings significantly expand results from a previous study that evaluated representation of women on NIH panels for a single year. These findings are limited by use of self-reported gender pronouns and the probability of a name belonging to a man or woman when pronouns were not reported. In addition, the random sampling may have introduced bias since some institutes and centers have more study sections than others.

## References

1. National Geographic Society. United States regions. Accessed June 5, 2022. <https://education.nationalgeographic.org/resource/united-states-regions>
2. Hansen M, Schoonover A, Skarica B, Harrod T, Bahr N, Guise JM. Implicit gender bias among US resident physicians. *BMC Med Educ.* 2019;19:396. doi:10.1186/s12909-019-1818-1
3. Newman EA, Waljee J, Dimick JB, Mulholland MW. Eliminating institutional barriers to career advancement for diverse faculty in academic surgery. *Ann Surg.* 2019;270(1):23-25. doi:10.1097/SLA.oooooooooooooo3273

<sup>1</sup>Baylor College of Medicine, Department of Internal Medicine, Houston, TX, USA, cole.wayant@gmail.com; <sup>2</sup>Oklahoma State University Center for Health Sciences, Department of Psychiatry and Behavioral Sciences, Tulsa, OK, USA

**Conflict of Interest Disclosures** Matt Vassar has received funding from the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism, the US Office of Research Integrity, Oklahoma Center for Advancement of Science and Technology, and internal grants from Oklahoma State University Center for Health Sciences.

## Global Gender Estimation From Distribution of First Names

Manolis Antonoyiannakis,<sup>1,2</sup> Hugues Chaté,<sup>3,4</sup> Serena Dalena,<sup>1</sup> Jessica Thomas,<sup>1</sup> Alessandro S. Villar<sup>1</sup>

**Objective** By construction, current methods of gender estimation portray gender-skewed populations as more gender-balanced than they truly are.<sup>1,2</sup> This systematic bias always underplays issues of underrepresentation, whereby one gender has a minority representation of less than 50%. A global method to estimate the gender composition of a population from correlations with first names was introduced that is free of systematic errors.<sup>3</sup> The method will improve our understanding of the review process and enhance analytics tools used in science.

**Design** Determining gender composition of a group from first names requires prior knowledge of name-gender correlations from a reference population. Current gender-estimation methods assume that name-gender conditional probabilities can be directly transferred from a reference population to a target population. This strong assumption means that one population must be a fair sample of the other, particularly in gender composition, implying that conventional methods will fail for strong gender asymmetry. A global gender estimator method (gGEM) was derived that instead quantifies how reference conditional probabilities must transform to best describe the observed list of names. The transformation, based on a process that morphs one population into another and seeks a self-consistent solution using the complete list of names, frees the estimation process from the fair-sampling assumption while also quantifying the strength of the otherwise hidden gender-dependent social process. Public data containing more than 200,000 names from 3 countries (40% from the US, 35% from Brazil, and

25% from France) were used as reference populations, from which prescribed fractions of men or women were removed to construct test populations of various gender compositions. The estimation method was compared with conventional approaches using these well-controlled test populations. A limitation is that the method is as accurate as the correlation between names and gender given by reference data.

**Results** gGEM provided accurate estimates irrespective of gender composition. It was observed that previous methods produced estimates that deviated linearly from the correct values as the gender mix deviated from gender balance. In the extreme case of a highly skewed test population composed of 1% women (correctly estimated by gGEM), previous methods estimated 3% to 2% prevalence of women depending on whether names with unclear gender were considered or not, respectively—a systematic error of at least 100% of the correct prevalence. gGEM showed no observable systematic effect for every gender mix tested. Typically, conventional methods incur systematic inaccuracy that grows quickly if the fraction of the underrepresented gender falls below 20 individuals per 100 people.

**Conclusions** When estimating the gender profile from first names, the global estimation method proposed here, which is easily implemented, should become the method of choice. Furthermore, it is argued that merging available reference populations with little overlap is a good strategy to mitigate errors stemming from population mismatching.

## References

1. Ross CO, Gupta A, Mehrabi N, Muric G, Lerman K. The leaky pipeline in physics publishing. *arXiv*. Preprint posted online October 18, 2020. doi:10.48550/arXiv.2010.08912
2. Squazzoni F, Bravo G, Farjam M, et al. Peer review and gender bias: a study on 145 scholarly journals. *Sci Adv*. 2021;7(2):eabdo299. doi:10.1126/sciadv.abdo299
3. gGEM. Home page. <https://www.ggem.app>

<sup>1</sup>American Physical Society, College Park, MD, USA, villar@aps.org; <sup>2</sup>Department of Applied Physics & Applied Mathematics, Columbia University, New York, NY, USA; <sup>3</sup>Service de Physique de l'Etat Condensé, CEA, CNRS, Université Paris-Saclay, CEA-Saclay, Gif-sur-Yvette, France; <sup>4</sup>Computational Science Research Center, Beijing, China

**Additional Information** Alessandro S. Villar and Hugues Chaté are co-corresponding authors.

## Analysis of Editorial Board Gender Parity of the Top 20 Most Influential Dermatology Journals

Mindy D. Szeto,<sup>1</sup> Torunn E. Sivesind,<sup>2</sup> Lori S. Kim,<sup>3</sup> Katie A. O'Connell,<sup>4</sup> Kathryn A. Sprague,<sup>5</sup> Yvonne Nong,<sup>6</sup> Daniel M. Strock,<sup>7</sup> Annie L. Cao,<sup>8</sup> Jieying Wu,<sup>9</sup> Lauren M. Toledo,<sup>10</sup> Sophia M. Wolfe,<sup>11</sup> Wyatt Boothby-Shoemaker,<sup>12</sup> Robert P. Dellavalle<sup>13</sup>

**Objective** Women continue to be underrepresented in academic leadership positions, especially in dermatology.<sup>1</sup>

Although women account for more than half of all board-certified dermatologists in the United States, academic dermatology leadership roles, such as department chair and fellowship director positions, remain disproportionately occupied by men.<sup>2</sup> Available evidence suggests that this inequity extends to medical journals, with substantial gender gaps reported in editorial board composition across multiple specialties. Previously published data from 2018 suggested that women accounted for the minority of dermatology editors in all positions.<sup>1</sup> To provide an update to this work and an evaluation of current trends, the gender distribution of dermatology editorial boards was assessed in 2021, making comparisons among the top 20 most influential dermatology journals.

**Design** The top 20 dermatology journals by 2020 *h*-index were identified on Scimago.<sup>3</sup> Journal editorial board websites were searched in November 2021 for lists of editor names and roles, and journal-defined editorial board members were collected and tabulated. Binary (women vs men) gender estimation by author first name was performed with Gender API, a popular gender inference service based on querying large multifactorial databases and name repositories. Estimations were corroborated by online searches of professional photographs and biographies by 2 independent reviewers, with in-depth discussion and consensus meetings to resolve discrepancies.

**Results** Women made up a mean (SD) of 37% (12%) and a median (IQR) of 33% (18%) of editorial boards across the top 20 dermatology journals (**Table 32**). The *Journal of Dermatological Science* (15%) and *Journal of the European Academy of Dermatology and Venereology* (22%) had the lowest proportion of women editors, while *Contact Dermatitis* (58%), *Sexually Transmitted Infections* (54%), and *Sexually Transmitted Diseases* (53%) had among the highest. The editorial board of *JAMA Dermatology* was observed to be 56% women after excluding International Advisory Committee members. Of the 20 journals, 5 journals (25%) had women editors in chief.

**Conclusions** This study found that underrepresentation of women on dermatology journal editorial boards persisted across multiple top journals. The finding suggests that editors in chief and journal leadership should consider establishing board member recruitment targets with the goal of gender parity. A target of 50% women on boards could more accurately represent the dermatology workforce. Limitations of our study include reliance on high-throughput software analyzing first names only and estimating binary gender, which may lead to misclassification. Future work should consider self-reported sex and gender identity to ensure true concordance with the individual's identity. Addressing gender gaps and encouraging diversity of identity and perspectives among editorial boards is a worthwhile goal for further research.

## References

1. Lobl M, Grinnell M, Higgins S, Yost K, Grimes P, Wysong A. Representation of women as editors in dermatology

**Table 32. Percentages of Women Editorial Board Members and Editors in Chief for the Top 20 Dermatology Journals by *h*-Index**

Dermatology journal	<i>h</i> -Index rank	<i>h</i> -Index 2020	Editorial board members, No.	Women, No. (%)	Woman editor in chief
<i>Journal of the American Academy of Dermatology</i>	1	208	209	79 (38)	No
<i>Journal of Investigative Dermatology</i>	2	201	51	21 (41)	No
<i>British Journal of Dermatology</i>	3	179	133	63 (47)	No
<i>JAMA Dermatology</i>	4	166	27	15 (56)	Yes
<i>Dermatologic Surgery</i>	5	125	122	38 (31)	No
<i>Lasers in Surgery and Medicine</i>	6	112	24	6 (25)	No
<i>Wound Repair and Regeneration</i>	7	109	61	26 (43)	No
<i>Journal of the European Academy of Dermatology and Venereology</i>	8	107	64	14 (22)	No
<i>Pigment Cell and Melanoma Research</i>	9	105	58	20 (34)	No
<i>Sexually Transmitted Diseases</i>	10	105	93	49 (53)	No
<i>Sexually Transmitted Infections</i>	11	98	82	44 (54)	Yes
<i>Contact Dermatitis</i>	12	96	36	21 (58)	Yes
<i>Experimental Dermatology</i>	13	96	143	44 (31)	No
<i>International Journal of Dermatology</i>	14	93	65	21 (32)	Yes
<i>Journal of Dermatological Science</i>	15	93	73	11 (15)	No
<i>Dermatology</i>	16	92	34	10 (29)	No
<i>American Journal of Clinical Dermatology</i>	17	89	40	10 (25)	Yes
<i>Clinics in Dermatology</i>	18	88	45	11 (24)	No
<i>Acta Dermato-Venereologica</i>	19	83	28	13 (46)	No
<i>Archives of Dermatological Research</i>	20	80	66	21 (32)	No
<b>Total</b>					
No. (%)	NA	NA	1454	537 (37)	5 (25)
Mean (SD), %	NA	NA	NA	37 (12)	NA
Median (IQR), %	NA	NA	NA	33 (18)	NA

Abbreviation: NA, not applicable.

journals: a comprehensive review. *Int J Womens Dermatol.* 2019;6(1):20-24. doi:10.1016/j.ijwd.2019.09.002

2. Nambudiri VE, Shi CR, Vleugels RA, Olbricht SM. Academic dermatology leadership in the United States—addressing the gender gap. *Int J Womens Dermatol.* 2018;4(4):236-237. doi:10.1016/j.ijwd.2018.05.003

3. SJR Scimago Journal & Country Rank. Accessed November 30, 2021. <https://www.scimagojr.com/journalrank.php>

<sup>1</sup>Department of Dermatology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA, mindy.d.szeto@gmail.com;

<sup>2</sup>Department of Dermatology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; <sup>3</sup>University of Illinois at Chicago College of Medicine, Chicago, IL, USA; <sup>4</sup>Eastern Virginia Medical School, Norfolk, VA, USA; <sup>5</sup>Chicago College of Osteopathic Medicine, Downers Grove, IL, USA; <sup>6</sup>Michigan State University College of Human Medicine, Flint, MI, USA; <sup>7</sup>Eastern Virginia Medical School, Norfolk, VA, USA; <sup>8</sup>Department of Dermatology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; <sup>9</sup>Rocky Vista University, Parker, CO, USA; <sup>10</sup>Lake Erie College of Osteopathic Medicine, Bradenton, FL, USA; <sup>11</sup>Department of Dermatology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; <sup>12</sup>Michigan State University College of Human Medicine, East Lansing, MI, USA; <sup>13</sup>Department of Dermatology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

**Conflict of Interest Disclosures** Robert P. Dellavalle is a joint coordinating editor for *Cochrane Skin*, a dermatology section editor for *UpToDate*, a social media editor for the *Journal of the American Academy of Dermatology*, a podcast editor for the *Journal of Investigative Dermatology*, editor in chief of *JMIR Dermatology*, and a coordinating editor representative on Cochrane Council. Torunn E. Sivesind serves as an editorial board member at large for *JMIR Dermatology* and receives fellowship funding from Pfizer. The remaining authors have no conflicts of interest to declare.

### Patient Involvement in CMAJ Publications From 2018-2020

Victoria Saigle,<sup>1,2</sup> Andreas Laupacis,<sup>1,3</sup> Kirsten Patrick<sup>1</sup>

**Objective** Despite increased integration of patient, family, and caregiver (henceforth referred to as “patient”) perspectives in health care and research, limited information exists regarding patient inclusion in academic publications,<sup>1</sup> and medical editors remain divided about the appropriateness of patient authorship.<sup>2</sup> Understanding the level of patient involvement can establish a baseline with which to assess the success of future efforts. This work examined the type and frequency of patient involvement in works published by a general medical journal prior to its adoption of a patient engagement policy.

**Design** In this cross-sectional study, articles (editorial, research, guideline, clinical review, commentary, humanities, analysis, and practice), podcasts, and blog posts published by the *CMAJ* between 2018 and 2020 were analyzed to assess how frequently patients were involved as authors, article contributors (eg, as members of advisory panels or patient-partnered teams), podcast interviewees, or subjects of patient profiles.

**Results** Overall, 97 of 973 articles (10.0%), 18 of 175 podcasts (10.3%), and 28 of 323 blog posts (8.7%) reflected patient involvement. The journal had published a special supplement focused on patient involvement in 2018 and launched a “Patient Portrait” series in 2020; these were the only instances in which the *CMAJ* editorial team explicitly solicited work involving patients. When articles published in special supplements were excluded, 6.6% of articles (62/933) involved patients. The highest rate of patient involvement was seen in guidelines (85% [17/20]), followed by commentaries (21.4% [37/173] when including articles published in special supplements and 1.5% [2/136] without), humanities (17.4% [34/195]) (which included “Patient Portraits”), analysis (6.3% [3/48]), research (2.4% [4/167]), and practice (0.6% [2/312]) articles. No editorials or clinical reviews involved patients in the period under study. Patient authorship was the most frequent type of involvement (46/97 [47.4%] overall; 25/34 [73.5%] humanities, 16/37 [43.2%] commentaries, 4/17 [23.5%] guidelines, and 1/2 [50.0%] practice articles). Other modes of patient involvement included patient-partnered teams (39/97 [40.2%]) or advisory groups (35/97 [36.1%]), incorporating patient preferences in methods (24/97 [24.7%]), patient profiles (8/97 [8.2%]), and mentioning patient contributions in the acknowledgment section (7/97 [7.2%]). Most patient authors of humanities articles were people who had worked in health care or research.

**Conclusions** At *CMAJ*, 10% of all published products (97/973) from 2018 to 2020 reflected patient (including family and caregiver) involvement. Overall, 6.6% (62/933) occurred without the journal explicitly soliciting content including patients. A large proportion of the articles that listed patient authors were published in special supplements or written by people who had worked in health care. These data will serve as a baseline as *CMAJ* seeks to increase patient involvement across the journal.

## References

1. Price A, Schroter S, Snow R, et al. Frequency of reporting on patient and public involvement (PPI) in research studies published in a general medical journal: a descriptive study. *BMJ Open*. 2018;8:e020452. doi:10.1136/bmjopen-2017-020452
2. Cobey KD, Monfaredi Z, Poole E, Proulx L, Fergusson D, Moher D. Editors-in-chief perceptions of patients as (co) authors on publications and the acceptability of ICMJE authorship criteria: a cross-sectional survey. *Res Involv Engagem*. 2021;7(1):39. doi:10.1186/s40900-021-00290-1

<sup>1</sup>*CMAJ*, Ottawa, Ontario, Canada, victoria.saigle@cmaj.ca; <sup>2</sup>Ottawa Hospital Research Institute, Ottawa, Ontario, Canada; <sup>3</sup>Dalla Lana School of Public Health and Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

**Conflict of Interest Disclosures** All authors are employed by the *CMAJ*.

## Editorial and Peer Review Process

### A Technology-Based, Quality Improvement Intervention to Ensure Accuracy and Integrity of the Scholarly Record of Articles Published Simultaneously in 2 Languages

Vivienne C. Bachelet,<sup>1,2,3</sup> Amaya Goyenechea,<sup>4</sup> Máximo Rousseau-Portalis<sup>5</sup>

**Objective** Academics in non–English speaking countries seek to publish in journals that promote locally generated knowledge in their language while appreciating having a broader reach to English-language audiences. Accordingly, a medical journal based in Latin America decided to become fully bilingual (Spanish and English). To achieve this objective within the constraints of a limited budget, it was necessary to modernize the editorial and publishing processes and platforms to ensure consistency and accuracy of both languages without overstretching existing resources. A strong editorial board commitment to bilingualism was required to avoid the simpler and less costly decision of publishing only in English.

**Design** A quality improvement intervention was designed to overhaul the journal’s peer review, copyediting, and publishing technologies to enable simultaneous publication of the original Spanish version with the journal-provided translation into English at no extra cost to authors, under a single digital object identifier. The intervention consisted of (1) designing a new peer-review software that applies an XML-first approach to the peer-review stage that allows full metadata capture at the start, and integrating multilingual options for both peer-review and copyediting, thus allowing handling of multilingual submissions seamlessly; (2) developing an open-source web-content publication platform to incorporate automated ingestion of XML, JATS, and PDF files and figures; and (3) modernizing the graphic interface and the article-level functionalities in the framework of a bilingual website. The intervention was partially funded by a 2-year grant from the Chilean Ministry of Science. The journal’s editor in chief and editorial staff were involved in all these stages as design partners.

**Results** Nine months after implementing the first component of the intervention, the outcomes were full bilingualism, shorter times from acceptance to publication of the article, and simultaneous publication of the Spanish and English versions (**Table 33**). The second and third components were in the final stages, and full deployment was expected by July 2022. Once 100% rollout of the bilingual copyediting software was achieved, built-in recycling of nontranslatable manuscript sections such as metadata,

**Table 33. Time Trend for Efficiency Indicators, 2019 to 2022**

Variable	2019	2020	2021	2022
<b>Bilingualism, by language, % of articles published</b>				
Spanish only	31	43	19	0
Spanish and English	62	45	69	94
English only <sup>a</sup>	7	12	12	6
<b>Time to publication, average, d</b>				
From acceptance to publication of Spanish	36	43	46	17
From publication of Spanish version to publication of English version	34	29	13	0

<sup>a</sup>English only refers to manuscripts submitted initially in English. The journal does not translate into Spanish.

keywords, abstracts, and references, would reduce inconsistencies between versions and unburden the human translators. Because deployment was still partial, the journal did offline technical editing of both languages simultaneously, thus ensuring the accuracy of both versions. The journal was financially sustainable despite the significant effort and rollout.

**Conclusions** Many resources were deployed in this intervention. All articles were published in both languages with a significantly shorter time to publication of the 2 languages. It is expected that, with time, artificial intelligence–based enhancements will result in greater automation and house styling of the Spanish and English versions of accepted manuscripts.

<sup>1</sup>Escuela de Medicina, Universidad de Santiago de Chile, Santiago, Chile; <sup>2</sup>Medwave; <sup>3</sup>Medwave Estudios Limitada, Santiago, Chile; <sup>4</sup>Medwave Estudios Limitada, Tel Aviv, Israel; <sup>5</sup>Laboratorio de Estadística Aplicada a las Ciencias de la Salud (LEACS), Departamento de Farmacología y Toxicología, Facultad de Medicina, Universidad de Buenos Aires, Buenos Aires, Argentina

**Conflict of Interest Disclosures** Vivienne C. Bachelet declared that she is the founding and managing partner of Medwave Estudios Limitada, the publishing company that owns the journal, and is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision of this abstract. No other conflicts were reported.

**Funding/Support** This project was cofunded by ANID (National Agency for Innovation and Development) of the Chilean Ministry of Science, Technology, Knowledge, and Innovation through the Scientific Journals Publication Fund, Call for Proposals 2020 (project code FP200001).

**Role of Funder/Sponsor** The funders had no role in the execution of this project other than oversight of the proper use of public funds. As the publishing company, Medwave Estudios is also financially responsible for the components of the project not covered by the public subsidy.

### Concordance Between Peer Reviewers' Recommendations and Editorial Decision-Making at *The Journal of Pediatrics*

Raye-Ann deRegnier,<sup>1,2</sup> Kevin Jewett,<sup>1,3</sup> Meghan McDevitt,<sup>4</sup> Denise M. Goodman<sup>1,2</sup>

**Objective** To examine concordance between peer reviewers' recommendations and editorial decisions for a general pediatric journal.

**Design** This was a cross-sectional observational study of a convenience sample of peer-review recommendations for articles submitted to *The Journal of Pediatrics* reported using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline. After submission, an editor made a preliminary assessment and invited peer reviewers. Responding reviewers were asked to provide a recommendation for acceptance, revision, or rejection. When recommending acceptance or revision, reviewers were asked to select a priority score for publication on a 4-level Likert scale (low, medium low, medium high, high). After assessing the reviews, editors decided either to move the article forward (accept or invite a revision) or to reject. Manuscripts were included in the study if the original submission was initially reviewed by at least 2 invited reviewers. The primary outcome of the study was the final disposition (accept or reject). Reviewer recommendations and Likert scores in response to the original submission were compared with final dispositions using univariable and multivariable regression. For the multivariable analysis, the highest reviewer priority score was dichotomized as high/medium high vs. medium low/low. Differences between priority scores were dichotomized as matched or differing by 1 Likert level vs differing by 2 or 3 Likert levels.

**Results** The sample included 1016 articles sent for review between June 16, 2020, and October 27, 2021, with at least 2 completed reviews in response to the original submission. At least 1 reviewer recommended rejection for 393 of 1016 manuscripts (38.7%). When reviewers agreed to reject, all manuscripts were rejected (65 of 65); however, when reviewers disagreed on rejection, 36 of 328 manuscripts were accepted for publication (11%;  $P = .005$ ) and 1 is pending to date. Among articles with a reviewer recommendation for rejection and at least 1 reviewer priority score ( $n=294$ ), high or medium high priority scores were associated with a final decision to accept the article (odds ratio [OR], 11.37; 95% CI, 3.94-32.85;  $P < .001$ ). Reviewers agreed to move the article forward for 623 of 1016 manuscripts (61.3%) with 395 ultimately accepted for publication. Of these 623, 494 (79.3%) had at least 2 reviewer priority scores and a final disposition. The highest reviewer priority score was significantly associated with the acceptance rate (Table 34). In multivariable regression, higher priority score category (OR, 12.59; 95% CI, 6.84-23.14) and level of reviewer agreement on priority scores (OR, 2.04; 95% CI, 1.15-3.6) were both significantly associated with article acceptance rates.

**Conclusions** Both reviewer recommendations on acceptance or rejection and priority scores for publication appeared to be influential in editorial decision-making. However, editors were frequently required to arbitrate mixed reviewer recommendations and make independent decisions.

**Table 34. Acceptance Rates by Reviewers' Priority Scores**

Highest priority score	Acceptance rate, % (No./total No.)				
	Acceptance rate (n = 494)	Priority scores match (n = 186)	1-Level Likert difference (n = 234)	2-Level Likert difference (n = 70)	3-Level Likert difference (n = 4)
High	90.4 (122/135)	93.8 (15/16)	97.1 (68/70)	80 (36/45)	75 (3/4)
Medium high	76.2 (224/294)	88.7 (110/124)	69.7 (101/145)	52 (13/25)	NA
Medium low	28.6 (18/63)	34.1 (29/44)	15.8 (3/19)	NA	NA
Low	0 (0/2)	0 (0/2)	NA	NA	NA

Abbreviation: NA, not applicable.

<sup>1</sup>The Journal of Pediatrics, r-deregnier@northwestern.edu; <sup>2</sup>Ann & Robert H. Lurie Children's Hospital of Chicago; <sup>3</sup>Cincinnati Children's Hospital; <sup>4</sup>Patient-Centered Outcomes Research Institute

**Conflict of Interest Disclosures** None reported.

### Trends in and Reasons for Peer Reviewers Declining Invitations to Review at *Diseases of the Colon & Rectum*, 2016-2021

Susan Galandiuk<sup>1</sup>

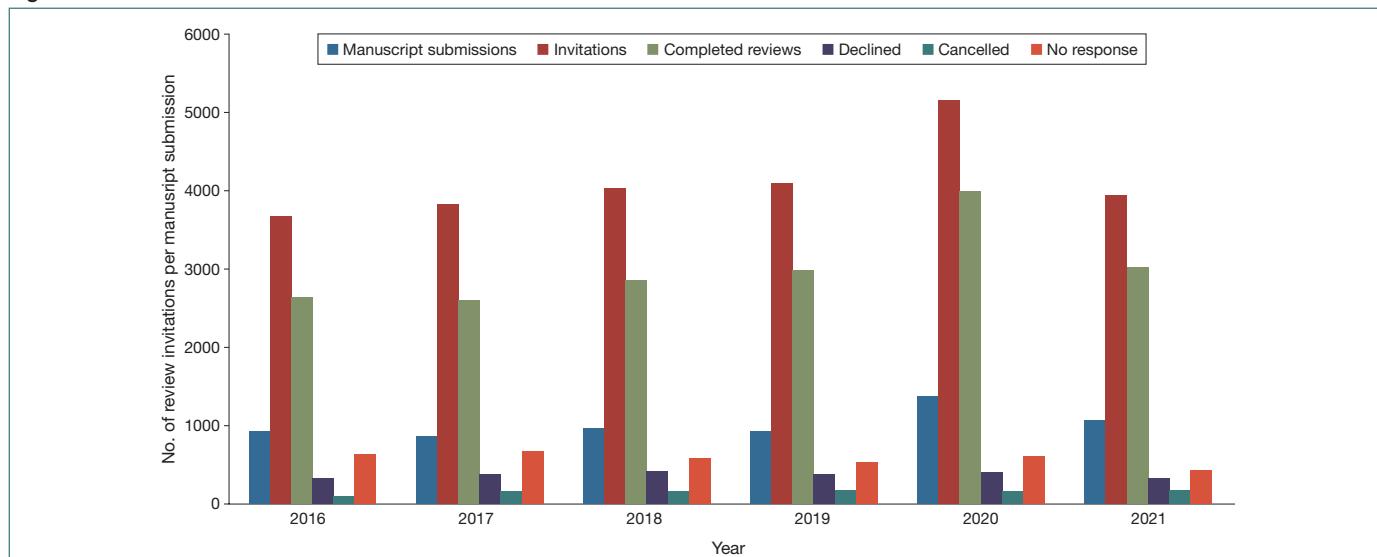
**Objective** Peer review is the backbone of scientific medical publishing. Most reviewers receive little reward and face competing demands for their time. *Diseases of the Colon & Rectum* is a surgical specialty journal that, similar to most journals, relies on peer review to define what is new and legitimate to publish. It was hypothesized that the frequency of “nonresponse” and “reviewer declines” responses to review requests increased over the last 6 years.

**Design** In this observational study, using 2 data sets, the editor in chief (EIC) (1) prospectively collected “reviewer declines” e-mail responses over a 6-year period, examining reasons for declining reviews, frequency of recommending alternate reviewers, and other comments and (2), for this

same 6 years, the EIC queried the manuscript submission and review system for total review invitations by editor, number declined, invitations without response, and for reviews cancelled due to noncompletion.

**Results** From January 2016 to December 2021, the journal received 872 to 1379 submissions per year (mean of 1024 submissions per year) for which 24,766 review invitations were sent, 17,306 by the EIC (70%). Overall, 19,004 invitations were accepted, of which 18,100 were completed on time (73%), 904 were initially accepted but cancelled due to noncompletion of the review (4%), 2235 invitations were declined (9%), and 3496 (14%) never received a response (**Figure 19**). The EIC prospectively collected 516 “reviewer-declines” e-mail responses from 474 unique reviewers representing a 23% sample of review declinations. Among these, 41% provided no reason and 34% were “busy with other commitments”; other common reasons included “away from the office” (8%), vacation (7%), and family obligations (2%), among others. Conflicts of interest were cited in 2% of declines and in less than 1% allegations of scientific misconduct were made by associate editors or reviewers for other journals. These cases involved the same manuscript and allegations of duplicate submission; the manuscript was withdrawn. Fourteen percent of declining reviewers suggested alternate reviewers (often with e-mail addresses). Reasons for declining did not change over time. Over the 6 years, EIC invitations became less likely to be declined and reviews cancelled for delay decreased. This appeared to coincide with development of a mentored peer-review program and institution of high-profile reviewer awards at the Society annual meeting. These data also encase the pandemic when editorial board members and reviewers were very involved in acute care of patients with COVID-19.

**Conclusions** This report becomes a baseline for ongoing quality improvement. There were no substantial changes in frequency of or reasons for declining reviews over time, particularly during the COVID-19 pandemic.

**Figure 19. Reviewer Data for All Editors**

<sup>1</sup>Hiram C. Polk, Jr. MD Department of Surgery, 550 South Jackson Street, University of Louisville, Louisville, KY, USA, sogala01@louisville.edu

**Conflict of Interest Disclosures** Susan Galandiuk receives a stipend for editorial work from the American Society of Colon & Rectal Surgeons.

## Ethics and Ethical Concerns

### Development and Testing of a Tool to Assist Editorial Staff in Review of Ethical Research Reporting in Manuscripts

Jan Higgins,<sup>1</sup> Robert D. Steiner,<sup>2</sup> Katharine Murphy,<sup>1</sup> Kyle Brothers<sup>3</sup>

**Objective** Journals are obligated to ensure published research is conducted ethically, with this task often falling to editorial staff. However, editorial staff are rarely trained in research ethics.

**Design** *Genetics in Medicine* (GIM) staff and editors developed a flowchart as a tool to categorize research studies submitted for publication. Twelve categories were defined (human participants or individual-level data; human biosamples/biobanks; transplantation; newborn screening blood spots; previously published data; database data; surveys; animal studies; electronic health records; quality assurance/quality control/operations; algorithm/software development; and deceased patient samples.) The GIM Ethics Advisory Committee (EAC) determined ethics statements required for publication. To test this tool, 50 consecutive articles from GIM prior to implementing the tool (2020) and 50 consecutive articles from 4 peer journals (*Human Medical Genetics*, *Genome Medicine*, *American Journal of Human Genetics*, and *npj Genomic Medicine* [all in 2020]) were analyzed. For each article, the following were assessed: (1) whether the flowchart provided appropriate categories; (2) whether an ethics statement was provided; and (3) whether the ethics statement met the standards defined by the GIM EAC.

**Results** Review of articles from GIM prior to implementation of the tool and from 4 peer journals revealed

that the majority comply with ethical standards (**Table 35**). A small number of articles lacked one or both elements of the ethics statement. These studies might have been conducted without appropriate permissions, or the journal might not have required a statement. The latter seems the most likely explanation for research using large databases, registries, or biobanks, as well as for studies based on previously published data. GIM now requires ethics statements for these categories, but other journals appear not to share this requirement.

**Conclusions** This tool facilitated review of articles by editorial office staff to ensure reporting of ethical conduct of research prior to publication and facilitates auditing for compliance with ethical standards.

<sup>1</sup>American College of Medical Genetics and Genomics, Bethesda, MD, USA, gim@acmg.net; <sup>2</sup>Marshfield Clinic Health System and University of Wisconsin School of Medicine and Public Health, Marshfield and Madison, WI, USA; <sup>3</sup>University of Louisville and Norton Children's Research Institute, Louisville, KY, USA

**Conflict of Interest Disclosures** Robert D. Steiner reports employment with PreventionGenetics/Exact Sciences; consulting and equity interest with Acer Therapeutics and PTC Therapeutics; consulting with Aeglea, Health Advances, Leadiant, Precision for Value, and Travere; honoraria from Medscape/WebMD and Teladoc; and research support from Alexion and the Smith Lemli Opitz Syndrome Foundation.

**Disclaimer** The content is solely the responsibility of the authors and does not necessarily represent the official views of the American College of Medical Genetics and Genomics.

### Research and Publication Ethics Knowledge and Practices in the Health and Life Sciences: Findings From an Exploratory Global Survey

Luchuo Engelbert Bain,<sup>1,2</sup> Ikenna Desmond Ebuenyi,<sup>3</sup> Jean Jacques Noubiap<sup>4</sup>

**Objective** To examine the levels of awareness, preferences, experiences, and practices of researchers in the health and life sciences regarding research and publication ethics.

**Design** In this cross-sectional study, a questionnaire was deployed on Google Forms to a global audience. Reminder

**Table 35. Results of the Ethics Statement Analysis**

Ethics statement meets or exceeds GIM expectations	Journal, No. of studies				
	<i>Human Medical Genetics</i>	<i>Genome Medicine</i>	<i>American Journal of Human Genetics</i>	<i>npj Genomic Medicine</i>	GIM
Yes	30	37	23	25	38
Yes for participant consent; no for research ethics committee review	1	0	1	0	0
No for participant consent; yes for research ethics committee review	1	0	1	0	0
No for both	9	8	1	1	5
Not provided, even though flowchart specifies statement is needed <sup>a</sup>	0	0	18	5	1
Not needed	9	5	6	4	6
Total	50	50	50	35 <sup>b</sup>	50

Abbreviation: GIM, *Genetics in Medicine*.

<sup>a</sup>Studies using large databases, registries, or biobanks and those using previously published data. These now require an ethics statement according to GIM standards.

<sup>b</sup>Only 35 studies were published in this journal in 2020.

emails and WhatsApp messages containing a brief description and a link to the online survey were sent through the various platforms until no new responses were received after 3 reminders. The form captured information regarding the sociodemographic characteristics of respondents. The questionnaire explored the level of awareness, training, attitudes toward, preferences, and experiences with research ethics committees. Regarding publication ethics, the awareness of respondents regarding the International Committee of Medical Journal Editors (ICMJE) authorship criteria and their experiences with ghost authorship were evaluated. All researchers working in the health and life sciences were eligible to participate in the study.

**Results** A total of 500 researchers were contacted, and 264 responded (53.0%). Only 36.7% of respondents were aware of the ICMJE authorship criteria. Less than a quarter (22.0%) of the respondents were aware of the existence of an ethics code. Respondents' experience with their most recent ethics approval application was poor (11.4%), good (36.0%), and excellent (4.2%). The practice of research teams to include authors with no or limited significant contribution to an article was frequent (44.3%), common (29.5%), and systemic (10.6%). More than 41.7% of the respondents had ever conducted a study involving human participant research without prior ethical approval. Respondents' experience with their most recent ethics approval application was poor (11.4%), good (36.0%), and excellent (4.2%). Major challenges in obtaining ethical approval were too much bureaucracy (47.3%), ethical approval application cost (5.3%), and unduly long review turnaround in receiving feedback and decisions (3.8%). Most respondents (83.0%) worked in institutions that hosted a research and ethics committee. Less than half (42.4%) of the respondents had been formally trained in publication ethics.

**Conclusions** Rates of ethical misconduct, such as having no ethical approval prior to conducting a study or gift and ghost authorships, were unacceptably high. Formal training in research and publication ethics should be institutionalized in the courses in universities and research institutions. Academic journals and funders have the duty to support researchers to uphold research ethics and research integrity standards. Institutionalization and awareness raising regarding these best practices are highly needed.

<sup>1</sup>Lincoln International Institute for Rural Health, College of Social Science, University of Lincoln. Brayford Pool, Lincoln, Lincolnshire, UK, lebain@gmail.com; <sup>2</sup>The Pan African Medical Journal, Yaoundé, Cameroon; <sup>3</sup>School of Nursing, Midwifery and Health Systems, University of Dublin, Ireland; <sup>4</sup>Centre for Heart Rhythm Disorders, University of Adelaide and Royal Adelaide Hospital, Adelaide, Australia

**Conflict of Interest Disclosures** None reported.

### Similarity Scores of Medical Research Manuscripts Before and After English-Language Editing

Joon Seo Lim,<sup>1</sup> Danielle A. Lee,<sup>1</sup> Sung-Han Kim,<sup>2</sup> Tae Won Kim<sup>3</sup>

**Objective** Plagiarism detectors are used by scientific journals to check submitted manuscripts for potential plagiarism. However, there are concerns that plagiarism detectors are overly sensitive and flag commonly used phrases as potential plagiarism.<sup>1</sup> This concern suggests that editing grammatically awkward phrases by using common expressions may increase similarity scores even though the flagged phrases were generated without intentions of plagiarism by the author or the editor. Considering that the use of language editing services is growing, this study compared the iThenticate similarity scores in medical research manuscripts before and after English-language editing.

**Design** This cross-sectional study was performed in January 2020 and is reported according to the STROBE guidelines. Fifty first-draft manuscripts written by the researchers at a tertiary referral center in Seoul, South Korea were randomly selected. All researchers were non-native-English-speaking Korean nationals. The similarity scores of the 50 manuscripts were assessed before and after English-language editing, which was provided by external vendors (40 [80%]) or in-house editors (10 [20%]). The default setting in iThenticate was used first according to the standard practice in scientific journals. Then, considering that the mean number of words per sentence in medical research papers is approximately 20 words,<sup>2</sup> the “exclude matches that are less than: 20 words” filter was also applied to exclusively assess sentence-level similarities. To assess the degree of text similarity according to each manuscript section, the similarity scores were manually measured by calculating the proportion of highlighted text in each section using ImageJ.

**Results** When using the default settings, the mean similarity scores of the 50 manuscripts increased from 28.5% to 30.2% after English-language editing, which was not statistically significant in an unpaired *t* test ( $P = .37$ ) but significant in a paired *t* test ( $P < .001$ ) (Table 36). Manually measured similarity scores of the manuscripts according to each manuscript section (Introduction, Methods, Results, and Discussion) also showed that the similarity scores of each manuscript significantly increased in a paired *t* test ( $P < .001$ ) but not in an unpaired *t* test ( $P = .07$ ); of the 4 sections, the Methods section had the highest mean similarity score both before and after English-language editing. When the “exclude matches that are less than: 20 words” filter was used to assess sentence-level similarity, the mean similarity score increased from 10.0% to 10.3%, which was not statistically significant in both unpaired ( $P = .85$ ) and paired ( $P = .57$ ) *t* tests.

**Conclusions** The similarity score of each manuscript showed a modest increase (mean change, 1.7%;  $P < .001$  in paired *t* test) when using the default setting of iThenticate. The default setting of iThenticate may lack the specificity to exclude inadvertent textual similarities, such as those resulting from the correction of grammatically incorrect phrases and the employment of common phrases in the course of English-language editing.

**Table 36. Text Similarity Scores Before and After English-Language Editing**

	Mean (SD)		<i>P</i> value (unpaired)	<i>P</i> value (paired)
	Before (n = 50)	After (n = 50)		
<b>Default setting</b>				
iThenticate score	28.5 (8.8)	30.2 (9.5)	.37	<.001
Manually measured similarity score, AU				
Total	17.1 (9.4)	19.3 (9.8)	.07	<.001
Introduction	17.4 (8.8)	19.2 (9.8)	.35	.005
Methods	26.7 (8.7)	29.2 (8.7)	.17	<.001
Results	12.2 (5.8)	14.5 (6.0)	.05	<.001
Discussion	12.0 (5.3)	14.2 (5.9)	.05	<.001
<b>20-Word setting (sentence-level similarity)</b>				
iThenticate score	10.0 (5.2)	10.3 (5.4)	.85	.57
Manually measured similarity score, AU				
Total	3.9 (3.1)	4.6 (3.6)	.34	.09
Introduction	2.2 (3.9)	3.0 (4.6)	.34	.07
Methods	10.4 (7.7)	11.4 (8.6)	.60	.22
Results	1.4 (2.7)	2.0 (3.1)	.45	.45
Discussion	1.7 (2.4)	1.9 (2.6)	.66	.77

Abbreviation: AU, arbitrary units.

## References

1. Weber-Wulff D. Plagiarism detectors are a crutch, and a problem. *Nature*. 2019;567(7749):435. doi:10.1038/d41586-019-00893-5
2. Barbic S, Durisko Z. Readability assessment of psychiatry journals. *Eur Sci Ed*. 2015;41:3-10.

<sup>a</sup>Scientific Publications Team, Clinical Research Center, Asan Institute for Life Sciences, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; <sup>b</sup>Department of Infectious Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; <sup>c</sup>Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea, twkimmd@amc.seoul.kr

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported by a grant (#2019-781) from the Asan Institute for Life Sciences at Asan Medical Center (Seoul, South Korea).

**Role of the Funder/Sponsor** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Funding/Sponsorship

### Comparison of Evaluations of Grant Proposals With and Without Numerical Scoring Submitted to Marie Skłodowska-Curie Actions' Innovative Training Networks

Ivan Buljan,<sup>1</sup> David G. Pina,<sup>2</sup> Antonia Mijatović,<sup>1</sup> Ana Marušić<sup>1</sup>

**Objective** The evaluation of European Union research grant proposals consists of 2 consecutive steps: (1) individual expert assessment and (2) consensus evaluation made by multiple reviewers. The result is an evaluation summary report, and previous studies have established this approach as a stable procedure in the assessment of research grants.<sup>1,2</sup> In 2020, numerical scores were replaced by textual comments in the individual expert assessment. The objective was to compare the linguistic characteristics of the comments for Excellence, Impact and Implementation criteria in evaluation reports of Marie Skłodowska-Curie Actions' Innovative Training Networks (ITN) proposals submitted in 2019 and 2020 to assess whether the removal of numerical scoring affected the structure of individuals evaluation report textual comments and evaluation outcome.

**Design** In this observational study for which data were collected in fall 2022, all ITN proposals submitted in 2019 and 2020 were considered. Information was collected about proposal scores and outcome, evaluation panel, and textual comments of the individual expert evaluations on all proposals submitted to the call. Linguistic characteristics of experts' comments were assessed using the Linguistic Inquiry and Word Count software, a program that counts words related to different psychological states and phenomena and gives a score that is a proportion of the specific category in the entire text. We used logistic regression to compare differences between the 2 call years, in which proposal variables (proposal status, word count for research excellence weaknesses, word count for implementation strengths, and negative effect levels for implementation strengths) were factors and the year of the call was criterion, with the significance level set at *P* < .001.

**Results** The number of proposals was similar in 2019 (*n* = 1554) and 2020 (*n* = 1503). The proportion of accepted proposals was slightly higher in 2020 (148 [9.85%]) than in 2019 (128 [8.24%]) (Table 37). In logistic regression,

**Table 37. Logistic Regression in Estimation of Horizon 2020 Innovative Training Networks 2020 Call Compared With Horizon 2020 Innovative Training Networks 2019 Call (*N* = 3057)<sup>a</sup>**

Variable	Level	Odds ratio	95% CI		Nagelkerke's <i>R</i> <sup>2</sup>
			Lower boundary	Upper boundary	
Intercept		0.42	0.33	0.55	
Proposal status	Ref (Rejected)				0.04
	Reserve	0.43	0.26	0.72	
	Accepted	1.33	1.02	1.72	
Word count for research excellence weaknesses (0-100)		1.01	1.00	1.01	
Word count for implementation strengths (0-100)		1.01	1.00	1.01	
Negative affect levels for implementation strengths (0-100)		0.64	0.50	0.81	

<sup>a</sup>Criterion variable was call year: 2019 call was labeled as 0 and 2020 call as 1.

experts' comments from 2020 differed from 2019 proposals in 2 linguistic domains. Comments in the Excellence section related to weaknesses had a greater number of words in the description of the proposal (**Table 37**). The comments on strengths in Implementation for proposals from 2020 had slightly more words and lower negative tone or words related to negative emotions, such as *wrong*, *suffer*, and *sad* (**Table 37**). All factors jointly explained around only 4% of the variance of the criterion.

**Conclusions** It seems that removing numerical scoring in the evaluation of ITN proposals at the stage of the individual assessment had little effect on the linguistic characteristics of the experts' comments, because all differences were marginal and we analyzed the whole proposal cohort.

## References

1. Pina DG, Buljan I, Hren D, Marušić A. A retrospective analysis of the peer review of more than 75,000 Marie Curie proposals between 2007 and 2018. *eLife*. 2021;10:e59338. doi:10.7554/eLife.59338
2. Buljan I, Pina DG, Marušić A. Ethics issues identified by applicants and ethics experts in Horizon 2020 grant proposals. *F1000Res*. 2021;10:471. doi:10.12688/f1000research.52965.2

<sup>1</sup>Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia, ibuljan@mefst.hr;

<sup>2</sup>European Research Executive Agency, European Commission, Brussels, Belgium

**Conflict of Interest Disclosures** David G. Pina is employed by the European Research Executive Agency. Ana Marušić occasionally serves as an ethics expert for the European Research Executive Agency and is an advisory board member of the Peer Review Congress but was not involved in the review or decision of this abstract. Ivan Buljan and Antonia Mijatović declare no competing interests.

**Funding/Support** This study was funded by the Croatian Science Foundation "Professionalism in Health – Decision making in practice and research" (ProDeM) under grant agreement IP-2019-04-4882.

**Role of the Funder/Sponsor** The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Disclaimer** All views expressed in this abstract are strictly those of the authors and may in no circumstances be regarded as an official position of the European Research Executive Agency or the European Commission.

## Misconduct

### Attitudes and Experiences of Authors, Reviewers, and Editors About Responsible and Detrimental Research Practices and the Transparency and Openness Promotion Guidelines Across Scholarly Disciplines

Mario Malički,<sup>1</sup> IJsbrand Jan Aalbersberg,<sup>2</sup> Lex Bouter,<sup>3,4</sup> Adrian Mulligan,<sup>5</sup> Gerben ter Riet<sup>1,6</sup>

**Objective** Transparency and Openness Promotion (TOP) guidelines were published in 2014, and since then, more than 5000 journals and 80 organizations have become TOP signatories.<sup>1</sup> However, attitudes toward all TOP guideline recommendations have not been systematically assessed. The goal of this study was to assess differences in attitudes and experiences of authors, reviewers, and editors about conducting and reporting research (TOP guidelines), their research environment, and perceived prevalence of responsible and detrimental research practices.

**Design** A survey was sent in 2018 to 100,000 randomly selected corresponding authors of articles indexed in Scopus, with 2 reminder rounds. The survey was designed for this study and consisted of 38 questions using 5-item rating scales (eg, from strongly agree to strongly disagree) and 10 (categorical) sociodemographic questions. Responses were presented as absolute numbers and percentages based on number of respondents per question. Ordinal regression analyses were used to explore associations between answers to questions and sociodemographic characteristics. All analyses were conducted in Stata, version 13, and  $P \leq .001$  was considered statistically significant.

**Results** Study response rate was 4.9% (3659 of 74,749 delivered emails) and included responses from 1389 authors, 1833 reviewers, and 434 editors. Respondents came from 126 countries, had a median age of 44 years (IQR, 35–55 years), and were most commonly from physical sciences (1034 [33%]) and life sciences (796 [25%]). There were no significant differences between authors, reviewers, and editors in their attitudes toward TOP guidelines, but some TOP recommendations (eg, study preregistration) were not supported. A majority (3462 [97%]) of respondents (strongly) agreed that researchers must appropriately cite study data, methods, and materials; 2675 (74%) that authors must follow appropriate reporting guidelines; 2174 (60%) that researchers must share data; 797 (23%) that authors must include the full data analysis plan in study preregistration; and 751 (21%) that studies must be preregistered. One fifth of respondents (701 [20%]) admitted sacrificing the quality of their publications for quantity, and 492 (14%) reported funders interfering in their study design or reporting. Undeserved authorship was perceived by all groups as the most prevalent detrimental research practice, while fabrication, falsification, plagiarism, and not citing prior relevant research were seen as more prevalent by editors than authors or reviewers (**Table 38**). Additional findings are available elsewhere.<sup>2</sup>

**Conclusions** There were no differences in attitudes between authors, reviewers, and editors toward specific TOP recommendations, although some recommendations were not supported. Respondents' perceptions of their research environments and of prevalence of detrimental research practices indicated that there is still much room for improvement. Without agreement and involvement of all stakeholders, it is unlikely that TOP recommendations will become standard practice. This study is limited by the overall

**Table 38. Respondents' Perceptions of Detrimental Research Practices**

Practice	Respondents, No. (%) <sup>a</sup>			
	Authors (n = 1184)	Reviewers (n = 1680)	Editors (n = 410)	Total (N = 3274)
Undeserved authorship (ie, guest or gift authorship)	406 (34)	674 (40)	160 (39)	1247 (38)
References omitted (ie, prior relevant research not cited)	314 (27)	585 (35)	190 (46)	1089 (33)
Ghost writing (ie, author(s) not acknowledged)	181 (15)	219 (13)	62 (15)	462 (14)
Undeclared conflict(s) of interest/competing interest(s)	194 (17)	200 (12)	58 (14)	452 (14)
Plagiarism	184 (16)	173 (10)	62 (15)	419 (13)
Fabrication or falsification	147 (12)	149 (9)	51 (12)	347 (11)

<sup>a</sup>Shown are number (percentage) of respondents who perceived the practices as (very) prevalent. Total number of respondents per question varied because the survey had no question that required a response.

low response rate, and therefore, the results might not be generalizable.

## References

- Nosek BA, Alter G, Banks GC, et al. Promoting an open research culture. *Science*. 2015;348(6242):1422-1425. doi:10.1126/science.aab2374
- Malički M, Aalbersberg IJ, Bouter L, Mulligan A, ter Riet G. Transparency in conducting and reporting research: a survey of authors, reviewers, and editors across scholarly disciplines. *Research Square*. Preprint posted online January 26, 2022. doi:10.21203/rs.3.rs-1296644/v1

<sup>1</sup>Urban Vitality Centre of Expertise, Amsterdam University of Applied Sciences, Amsterdam, the Netherlands, mario.malicki@mefst.hr; <sup>2</sup>Elsevier, Amsterdam, the Netherlands; <sup>3</sup>Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, the Netherlands; <sup>4</sup>Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, the Netherlands; <sup>5</sup>Elsevier, Oxford, UK; <sup>6</sup>Department of Cardiology, Amsterdam University Medical Centers, Amsterdam, the Netherlands

**Conflict of Interest Disclosures** Mario Malički is co-Editor in Chief of *Research Integrity and Peer Review*. IJsbrand Jan Aalbersberg is Senior Vice President of Research Integrity at Elsevier. Lex Bouter is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. Adrian Mulligan is a research director for Customer Insights at Elsevier. No other disclosures were reported.

**Funding/Support** Elsevier funded the work of Mario Malički in this study.

## Degree of Text Similarity and Prevalence of Potential Plagiarism in Biomedical Research Articles According to Linguistic Background and Field of Study

Joon Seo Lim,<sup>1</sup> Danielle A. Lee,<sup>1</sup> Sung-Han Kim,<sup>2</sup> Tae Won Kim<sup>3</sup>

**Objective** Text similarity detection software is widely used by biomedical journals to screen submitted manuscripts for potential plagiarism, with some journals rejecting manuscripts with high overall similarity scores in (eg, >40%) without further review. However, considering that overall scores may be vulnerable to false-positives resulting from common phrases, certain guidelines suggest examining the single-source scores to detect potential plagiarism.<sup>1</sup> The degree of text similarity and prevalence of potential plagiarism in biomedical articles was examined according to linguistic background (English-speaking vs non-English-speaking) and field of study (clinical vs nonclinical).

**Design** This cross-sectional study was performed in June 2020 and followed the STROBE reporting guideline. We analyzed the iTThenticate similarity reports of 480 articles randomly selected from an open access multidisciplinary journal, *PLoS One*. The articles were categorized into 8 preselected countries as English-speaking (USA, UK, Canada, Australia) vs non-English-speaking (Korea, China, France, Italy) and 6 fields of study as clinical (cardiology, gastroenterology, oncology) vs nonclinical (molecular biology, genetics, microbiology). The degree of text similarity was defined as the overall iTThenticate score, and the presence of potential plagiarism was defined as either (1) a single-source score of greater than 10% according to the Springer Nature guideline<sup>1</sup> or (2) overall score of greater than 40%, which is a cutoff used at some journals for considering editorial actions.<sup>2,3</sup> The similarity scores in each manuscript section were measured by calculating the proportion of highlighted text in each using ImageJ.

**Results** The degree of text similarity differed significantly among countries, with articles from non-English-speaking countries having higher scores than those from English-speaking countries (30.9% vs 23.8%, respectively;  $P < .001$ ) (**Table 39**). Among the non-English-speaking countries, there was no significant difference in the degree of text similarity between Asian and European countries (31.7% vs 30.1%, respectively;  $P = .27$ ). Text similarity also differed among fields of study, with clinical articles having higher scores than nonclinical articles (29.5% vs 25.2%, respectively;  $P < .001$ ). Measurement of text similarity showed that the Methods had the highest degree of text similarity among manuscript sections. The overall prevalence of potential plagiarism was 13.5% (65/480) and 13.8% (66/480) according to the single-source score cutoff of greater than 10% and the overall score cutoff of greater than 40%, respectively. Except for the lower prevalence of potential plagiarism in English-speaking countries according to the overall score cutoff (5.4% vs 22.1%, respectively;  $P < .001$ ), no statistically significant differences were noted between English-speaking and non-English-speaking countries, Asian and European countries, and clinical and nonclinical articles.

**Conclusions** While the degree of text similarity differed significantly according to linguistic background and field of study, the prevalence of potential plagiarism was similar across countries and fields of study. Clinical researchers in

**Table 39. Degree of Text Similarity and Prevalence of Potential Plagiarism According to Linguistic Background, Region, and Field of Study**

	Linguistic background			Region			Field of study		
	English-speaking (n = 240)	Non-English-speaking (n = 240)	P value	Asia (n = 120)	Europe (n = 120)	P value	Clinical (n = 240)	Nonclinical (n = 240)	P value
iThenticate similarity score, mean (SD), %	23.8 (9.6)	30.9 (10.8)	<.001	31.7 (11.5)	30.1 (10.1)	.27	29.5 (10.2)	25.2 (11.1)	<.001
<b>Manually measured similarity score, mean (SD), AU</b>									
Introduction	32.0 (15.5)	36.5 (16.8)	.06	35.1 (15.6)	38.3 (17.2)	.76	35.6 (15.9)	32.8 (16.0)	.08
Methods	38.3 (15.8)	43.3 (15.7)	.03	43.7 (16.3)	43.7 (15.1)	>.99	41.9 (15.6)	38.8 (15.8)	.15
Results	21.9 (7.8)	27.3 (9.3)	<.001	30.1 (10.6)	26.4 (8.8)	.14	27.3 (9.3)	22.6 (8.8)	<.001
Discussion	22.8 (8.8)	29.6 (10.0)	<.001	31.7 (11.3)	29.5 (10.1)	.99	28.5 (9.9)	25.2 (10.8)	.004
<b>Prevalence of potential plagiarism, No./total No. (%)</b>									
Single-source score >10%	30/240 (12.5)	35/240 (14.6)	.48	18/120 (15.0)	17/120 (14.2)	.70	40/240 (16.7)	25/240 (10.4)	.07
Overall score >40%	13/240 (5.4)	53/240 (22.1)	<.001	27/120 (22.5)	26/120 (21.7)	.88	38/240 (15.8)	28/240 (11.7)	.19

Abbreviation: AU, arbitrary units.

non-English-speaking countries in particular may benefit from receiving English-language writing education to avoid unintended text similarity.

## References

1. Springer. Plagiarism prevention with CrossCheck. Accessed February 24, 2022. <https://www.springer.com/gp/authors-editors/editors/plagiarism-prevention-with-crosscheck/4238>
2. IEEE Robotics & Automation Society. Information for IROS editors. Accessed June 14, 2022. <https://www.ieee-ras.org/conferences-workshops/financially-co-sponsored/iros/information-for-editors>
3. ARRUS Journal of Mathematics and Applied Science. Plagiarism policy. Accessed June 14, 2022. <https://jurnal.ahmar.id/index.php/mathscience/plagiarism>

<sup>1</sup>Scientific Publications Team, Clinical Research Center, Asan Institute for Life Sciences, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; <sup>2</sup>Department of Infectious Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; <sup>3</sup>Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; twkimmd@amc.seoul.kr

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported by grant 2019-781 from the Asan Institute for Life Sciences at Asan Medical Center, Seoul, South Korea.

**Role of the Funder/Sponsor** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Open and Public Access

### European Scholarly Journals From Small and Mid-Size Publishers in Times of Open Access: Mapping Journals and Public Funding Mechanisms

Mikael Laakso,<sup>1</sup> Anna-Maija Multas<sup>2</sup>

**Objective** This study investigated the relationship between government funding and scholarly journal publishing, specifically concerning small and mid-sized journal publishers in European countries. This study was conducted against the changing backdrop of the global scholarly journal landscape in which an increasing share of journals are free for both authors and readers, which raises the need for journals to seek resources other than subscription-based income.<sup>1,2</sup>

**Design** To achieve this objective, the study included 2 components: (1) establish the volume and key bibliometric characteristics of small and mid-sized journal publishers (an organization or actor publishing ≤150 peer-reviewed journals) present in European countries and (2) collect information about country-level public funding mechanisms for scholarly journals active in the 51 sovereign states in Europe, including transcontinental states partly in Europe,<sup>3</sup> as well as Kazakhstan and Kosovo. The Ulrichsweb publications database was used for the bibliometric component of the study. For funding information, manual data collection was required to gather as much information as possible, which included explorative web searches and consultations with scholarly publishing experts in the countries of interest.

**Results** The study identified 16,387 journals from small and mid-sized publishers being published in European countries, of which 36% were already publishing open access, ie, free for anyone to read on the web. **Table 40** presents a geographic breakdown of the journal counts and open access status per publisher size category. The majority of journals published in Europe were published by single-journal publishers (77% of all publishers), ie, by actors or organizations that only output 1 journal. Journals from small and mid-sized publishers were found to be multilingual or non-English to a higher degree than larger publishers that publish in excess of 150 journals each (44% and 43% vs 6% and 5%, respectively). Substantial diversity was observed in how (and whether) countries reserve and distribute funds to journals active in those countries, ranging from continuous inclusive subsidies to

**Table 40. Geographic Breakdown of Journal Counts and Open Access (OA) Status per Publisher Size Category**

Region	Total No. of journals	Large publishers					Small and mid-sized publishers				
		No. of journals	% of All journals	No. of subscription journals	No. of OA journals	% of Large-publisher journals that are OA	No. of journals	% of All journals	No. of subscription journals	No. of OA journals	% of Small- and mid-sized-publisher journals that are OA
Northern Europe	815	47	6	27	20	43	768	94	380	388	51
Eastern and Central Europe	7985	301	4	92	209	69	7684	96	5100	2584	34
Southern Europe	3167	125	4	94	31	25	3042	96	1428	1614	53
Western Europe	14,610	9717	67	7696	2021	21	4893	33	3585	1308	27
Total	26,577	10,190	38	7909	2281	22	16,387	62	10,493	5894	36

competitive grant funding to nothing at all. Information about funding instruments was overall hard to locate and often only available in national languages.

**Conclusions** Overall, the study suggests that the European journal publishing and funding space for small and mid-sized publishers is very diverse when it comes to whether and how economic and technical support from government resources is offered. Many countries have recently set up journal portals for hosting open access journals published in these countries, and funding instruments often require funded journals to enable open access for their content. Funding information was often difficult to discover, and efforts to make such information more easily available would likely facilitate policy development in this area.

## References

1. Bosman J, Frantsvåg JE, Kramer B, Langlais P-C, Proudman V. OA diamond journals study, part 1: findings. *Zenodo*. Published online March 9, 2021. doi10.5281/zenodo.4558704
2. Crawford W. Gold Open Access 2016-2021: Articles in Journals (GOA7). Cites & Insights Books; 2022.
3. EUR-Lex. Browse by EuroVoc. 2021. Accessed June 9, 2022. [https://eur-lex.europa.eu/browse/eurovoc.html?params=72#arrow\\_7206](https://eur-lex.europa.eu/browse/eurovoc.html?params=72#arrow_7206)

<sup>1</sup>Hanken School of Economics, Information Systems Science, Helsinki, Finland, mikael.laakso@hanken.fi; <sup>2</sup>Information Studies, Faculty of Humanities, University of Oulu, Oulu, Finland

**Conflict of Interest Disclosures** Mikael Laakso is the chair of the Finnish Association for Scholarly Publishing.

**Funding/Support** This research was funded by the Finnish Association for Scholarly Publishing.

**Role of the Funder/Sponsor** The funder had no influence on the execution of the study.

**Acknowledgment** The authors wish to express their gratitude to the journal editors who took the time to respond to our survey, and

to the OpenAIRE contact persons for giving helpful information on the public funding circumstances in many of the included countries.

## Open Access and Copyright License Status of Pharmaceutical Company-Supported Articles

Elin Bevan,<sup>1</sup> Tim Koder,<sup>1</sup> Valérie Philippon,<sup>2</sup> Slávka Baróniková,<sup>3</sup> Larisa Miller,<sup>4</sup> William Gattrell,<sup>5</sup> Tomas Rees<sup>1</sup>

**Objective** Informatics approaches have previously been used to assess open access (OA) rates for pharmaceutical company-supported publications.<sup>1</sup> An updated method was developed using article type metadata supplied by Embase, aiming to improve differentiation between journal articles and published congress materials. Copyright license types of OA pharmaceutical company-supported journal articles were also identified.

**Design** Articles from 24 pharmaceutical companies were analyzed, including the top 20 companies by revenue (as of September 2021)<sup>2</sup> and the 4 OpenPharma participant companies outside the top 20.<sup>3</sup> A scholarly search aggregator (Lens.org) was used to identify articles with authors affiliated with each company, a field of study of medicine, and publication in 2019 or 2020. Only publications tagged as article by Embase were included. The OA status and copyright license type were evaluated through Unpaywall (a database of OA articles). The 2019 analysis was performed using the original method based on PlumX-supplied data to establish any differences between the 2 methods. Data collection occurred from July to September 2021.

**Results** For 2019, the updated method returned fewer total articles (5093) than the original method (6900), with an overall OA rate of 76%, an increase from 69%. Analysis of 1 company with a known 100% OA rate revealed that in 2019, using the previous methods, 15 of 50 articles were classified incorrectly as non-OA; with the new method for the same data set, the number of articles was 2 of 37. The 2020 OA rate using the revised method was 77%, from a total of 5678 articles. Excluding companies with fewer than 10 articles (2019, 2 of 24; 2020, 1 of 24), the range of company OA rates

was 67% to 95% in 2019 and 56% to 90% in 2020. OpenPharma company participants had higher OA rates than nonparticipants, and their OA rates increased from 2019 to 2020 by 0.5 percentage point (77.1% to 77.6%) compared with 1.5 percentage points for nonparticipants (74.6% to 76.1%). The most common copyright license types in the 8242 OA journal articles published in 2019 and 2020 combined were CC BY-NC-ND (29% of total articles) and CC BY (28% of articles) (**Table 41**). Unpaywall was unable to distinguish the copyright license type for 21% of OA articles.

**Conclusions** The revised method improved accuracy of estimating OA rates for pharmaceutical company-supported articles by excluding incorrectly tagged publications identified with the original method, likely because Embase includes a manual step in their tagging. Almost half of the articles analyzed had a copyright license preventing commercial use. This analysis was restricted to only those articles with pharmaceutical company authors and articles tagged as “medicine”; therefore, this analysis does not encompass the full range of publications associated with the pharmaceutical industry. However, analyses using this method could help pharmaceutical companies, collaborating academic institutions, and publishers working toward greater openness and expanded reach of peer-reviewed publications.

## References

1. Freeman H, Macdonald S, Baróniková S, et al. Original abstracts from the 2021 European Meeting of ISMPP. *Curr Med Res and Opin.* 2021;37:21-39. doi:10.1080/03007995.2021.1894669
2. Fiercepharma.com. The top 20 pharma companies by 2020 revenue. Accessed January 28, 2022. <https://www.fiercepharma.com/special-report/top-20-pharma-companies-by-2020-revenue>
3. OpenPharma. About us. Accessed January 28, 2022. <https://openpharma.blog/about-us/>

<sup>1</sup>Oxford PharmaGenesis, Oxford, UK, elin.bevan@pharmagenesis.com; <sup>2</sup>Takeda Development Center Americas, Inc, Cambridge, MA, USA; <sup>3</sup>Galápagos NV, Mechelen, Belgium; <sup>4</sup>Alexion, AstraZeneca Rare Disease, Boston, MA, USA; <sup>5</sup>Ipsen, Abingdon, UK

**Conflict of Interest Disclosures** Elin Bevan, Tim Koder, and Tomas Rees are employees of Oxford PharmaGenesis. Slávka

**Table 41. OA License Types of Pharmaceutical Company-Supported OA Articles in 2019 and 2020**

OA license type	Articles, No. (%)
CC BY	2321 (28)
CC BY-NC	1606 (19)
CC BY-NC-ND	2400 (29)
No data	1692 (21)
Other <sup>a</sup>	223 (3)

Abbreviations: CC BY, attribution; CC BY-NC, attribution-noncommercial; CC BY-NC-ND, attribution-noncommercial-no derivs; OA, open access.

<sup>a</sup>Author choice/editor's choice usage agreement: Creative Commons, attribution noncommercial ShareAlike (CC BY-NC-SA), attribution ShareAlike (CC BY-SA), CC BY-ND; Elsevier-specific: OA user license, implied OA, public domain, publisher-specific license (as categorized by Unpaywall).

Baróniková is an employee of and may hold stock in Galápagos NV. Larisa Miller is an employee and stockholder of Alexion, AstraZeneca Rare Disease. At the time of abstract development, William Gattrell was an employee and shareholder of or held stock or stock options in Ipsen and is now an employee of Bristol Myers Squibb; Valérie Philippon was an employee and shareholder of or held stock or stock options in Takeda Development Center Americas, Inc, and is now an employee of UCB.

**Funding/Support** This work was produced by representatives of OpenPharma. OpenPharma is a multisponsor collaboration facilitated by Oxford PharmaGenesis Ltd and receives sponsorship funding from Alexion Pharmaceuticals Inc, AstraZeneca Pharmaceuticals LP, Boehringer Ingelheim International GmbH, E. R. Squibb & Sons LLC, F. Hoffmann-La Roche AG, Galápagos NV, Gilead Sciences Inc, GlaxoSmithKline Biologicals SA, Ipsen Biopharm Ltd, Janssen Global Services LLC, John Wiley & Sons Ltd, Novartis Pharma AG, Novo Nordisk A/S, Pfizer Inc, Takeda Development Center Americas Inc, and UCB Biopharma SRL.

**Role of the Funder/Sponsor** The funding organizations involved in this work facilitated the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the abstract; and were involved in the decision to submit the abstract for presentation.

**Acknowledgment** We thank the members and supporters of Open Pharma for their input and insightful discussions surrounding this analysis.

## Open Science

### Proportion of Academic Institutions With Courses on Open and Reproducible Science and Characteristics of the Courses

Hassan Khan,<sup>1,3</sup> Mona Ghannad,<sup>1,9</sup> Elham Almoli,<sup>1,4</sup> Marina Christ Franco,<sup>1,5</sup> Jeremy Ng,<sup>1</sup> Ana Patricia Ayala,<sup>6</sup> Emma L. Henderson,<sup>7</sup> Clare Ardern,<sup>8</sup> Kelly Cobey,<sup>9</sup> Sara Saba,<sup>10</sup> David Moher<sup>1,2</sup>

**Objective** Previous research has shown that academic institutions, in general, have neither endorsed nor implemented open science (OS) practices among faculty when it comes to promotion, tenure, and hiring.<sup>1,2</sup> Academic institutions play an integral role in driving a culture shift toward OS. Through raising awareness of the potential benefits and pitfalls of OS and providing necessary skills and training to their learners, staff and faculty are crucial in facilitating a culture change.<sup>3</sup> There is limited research examining the degree to which academic institutions prepare their students and trainees for best practices and/or government mandates in OS. The aim of this study was to examine (1) the proportion of a subset of academic institutions currently teaching a course on OS and (2) the characteristics of the eligible courses based on a set criteria of 6 OS topics.

**Design** This cross-sectional study examined the teaching of OS courses from January 2015 onward at the undergraduate and graduate level in a random global sample of 127 academic institutions. Academic institutions were selected based on the Centre for Science and Technology Leiden 2021 world ranking based on the proportion of open access publications.

Courses with at least 6 consecutive and thematic lectures with a university or equivalent departmental code were eligible and were assessed on 6 OS topics (reproducibility [crisis] and/or replication; design, methods, or data [code] material transparency; registration and/or preregistration; publishing of research and publication models; conceptual and statistical knowledge; and academic life and culture) based on the Framework for Open and Reproducible Research Training and the Transparency and Openness Promotion guidelines.

**Results** Of the 127 academic institutions examined, 65 (51%) had accessible course catalogs and/or course descriptions. Fifty-four institutions (83%) were identified as having previously taught or currently teaching a course or courses on OS. Overall, 72 possible OS courses were identified, with 4 OS course syllabi (6%) currently accessible. Three of the syllabi discussed reproducibility (crisis)/replicability; design, methods, or data (code) transparency; and conceptual and statistical knowledge. Two discussed academic culture and registration (and/or preregistration), although neither discussed publishing of research and publication models.

**Conclusions** This study provides a snapshot of the proportion of a subset of academic institutions currently teaching a course on OS and the depth of OS topics being taught to learners. This study highlights the extent to which academic institutions are fostering a learning environment that supports OS in higher education.

## References

1. Rice DB, Raffoul H, Ioannidis JPA, Moher D. Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities. *BMJ*. 2020;369:m2081. doi:10.1136/bmj.m2081
2. Khan H, Almoli E, Franco MC, Moher D. Open science failed to penetrate academic hiring practices: a cross-sectional study. *J Clin Epidemiol*. 2022;144:136-143. doi:10.1016/j.jclinepi.2021.12.003
3. League of European Research Universities. Open science and its role in universities: a roadmap for cultural change. Published May 2018. Accessed January 24, 2021. <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>

<sup>1</sup>Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, hassankhan@cmail.carleton.ca; <sup>2</sup>School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada; <sup>3</sup>Faculty of Arts and Social Sciences, Carleton University, Ottawa, Ontario, Canada; <sup>4</sup>School of Interdisciplinary Science, McMaster University, Hamilton, Ontario, Canada; <sup>5</sup>School of Dentistry, Federal University of Pelotas, Pelotas, Brazil; <sup>6</sup>Gerstein Science Information Centre, University of Toronto, Toronto, Ontario, Canada; <sup>7</sup>School of Psychology, University of Surrey, Surrey, England; <sup>8</sup>Faculty of Medicine, University of British Columbia, British Columbia, Canada; <sup>9</sup>University of Ottawa Heart Institute, Ottawa, Ontario, Canada; <sup>10</sup>Faculty of Science, University of Ottawa, Ottawa, Ontario, Canada

**Conflict of Interest Disclosures** There are no conflicts of interest to declare related to this study. Some of the members of the study are involved with initiatives related to research transparency,

including the Hong Kong Principles, The Centre for Open Science, and Declaration on Research Assessment. David Moher is an associate director of the International Congress on Peer Review and Scientific Publication but was not involved in the review or decision of this abstract.

## Pandemic Science

### Agreement of Treatment Effect Estimates From Observational Studies and Randomized Clinical Trials Evaluating Therapeutics for COVID-19

Osman Moneer,<sup>1</sup> Garrison Daly,<sup>2</sup> Joshua J. Skydel,<sup>3</sup> Kate Nyhan,<sup>4</sup> Peter Lurie,<sup>2</sup> Joseph S. Ross,<sup>5,6,7</sup> Joshua D. Wallach<sup>8</sup>

**Objective** To systematically identify, match, and compare treatment effect estimates and study demographic characteristics from observational studies and randomized clinical trials (RCTs) evaluating the same COVID-19 therapeutics, comparators, and outcomes.

**Design** In this meta-epidemiological study, individual RCTs or meta-analyses of RCTs reported in a *BMJ* living review directly comparing any of the 3 most frequently studied therapeutic interventions for COVID-19 (hydroxychloroquine-chloroquine, lopinavir-ritonavir, or dexamethasone) were identified for any safety and efficacy outcomes.<sup>1</sup> Using the Epistemonikos “Living Overview of Evidence” evidence database, individual observational studies evaluating the same interventions, comparisons, and outcomes reported in the *BMJ* review were identified. Treatment effect estimates from observational studies were identified, standardized, and, when possible, meta-analyzed to match individual RCTs or meta-analyses of RCTs with the same interventions, comparisons, and outcomes (ie, matched pairs). The direction and statistical significance (both  $P < .05$  or  $P \geq .05$ ) of treatment effect estimates and the distribution of study demographic characteristics from matched pairs were then compared.

**Results** Seventeen new, independent meta-analyses of observational studies were conducted of hydroxychloroquine-chloroquine, lopinavir-ritonavir, or dexamethasone vs an active or placebo comparator for any safety or efficacy outcomes and were matched and compared with 17 meta-analyses of RCTs reported in the *BMJ* review. Ten additional matched pairs with only 1 observational study and/or only 1 RCT were identified. Across all 27 matched pairs, 22 included any demographic and clinical data for all individual studies. All 22 matched pairs had studies with overlapping distributions of sex, age, and disease severity. Overall, 21 (78%) of the 27 matched pairs had effect estimates that agreed in terms of direction and statistical significance (**Table 42**). Higher levels of concordance were observed among the 17 matched pairs consisting of meta-analyses of observational studies and meta-analyses of RCTs (14 [82%]) than among the 10 matched pairs consisting of only 1 observational study and/or only 1 RCT (7 [70%]). The 18 matched pairs with relative treatment effect estimates also had higher levels of agreement (16 [89%]) than the 9 matched

**Table 42. Concordance Between Treatment Effect Estimates From 27 Matched Observational Study and RCT Pairs**

Observational study treatment effect estimates	RCT treatment effect estimates			
	Increased, statistically significantly	Decreased, statistically significantly	Increased, but not statistically significantly	Decreased, but not statistically significantly
Matched pairs consisting of meta-analyses of observational studies and meta-analyses of RCTs				
Increased, significantly <sup>a</sup>	0 <sup>b</sup>	0	2	0
Decreased, significantly <sup>a</sup>	0	0 <sup>b</sup>	0	0
Increased, but not significantly <sup>c</sup>	0	0	4 <sup>b</sup>	2 <sup>b</sup>
Decreased, but not significantly <sup>c</sup>	0	1	5 <sup>b</sup>	3 <sup>b</sup>
Additional matched pairs consisting of 1 observational study and/or 1 RCT				
Increased, significantly <sup>a</sup>	0 <sup>b</sup>	0	0	1
Decreased, significantly <sup>a</sup>	0	0 <sup>b</sup>	0	1
Increased, but not significantly <sup>c</sup>	1	0	3 <sup>b</sup>	1 <sup>b</sup>
Decreased, but not significantly <sup>c</sup>	0	0	1 <sup>b</sup>	2 <sup>b</sup>

Abbreviation: RCT, randomized clinical trial.

<sup>a</sup>Statistically significant based on  $P < .05$ .

<sup>b</sup>Pairs classified as concordant.

<sup>c</sup>Not statistically significant based on  $P \geq .05$ .

pairs with continuous treatment effect estimates (5 [56%]). Although 37 (80%) of the 46 individual observational studies referenced at least 1 RCT, only 12 (32%) of the 37 relevant RCTs were referenced by at least 1 observational study.

**Conclusions** More than three-quarters of the matched pairs had treatment effects that were in agreement. Meta-analyses of observational studies and RCTs evaluating therapeutics for the treatment of COVID-19 more often than not have summary treatment effect estimates that are in agreement in terms of direction and statistical significance. Although concerns have been raised about the evidence produced by individual observational studies evaluating therapeutics for COVID-19,<sup>2</sup> meta-analyzed evidence from observational studies may complement evidence collected from RCTs.

## References

- Siemieniuk RA, Bartoszko JJ, Ge L, et al. Drug treatments for COVID-19: living systematic review and network meta-analysis. *BMJ*. 2020;370:m2980. doi:10.1136/bmj.m2980
- Jung RG, Di Santo P, Clifford C, et al. Methodological quality of COVID-19 clinical research. *Nat Commun*. 2021;12(1):943. doi:10.1038/s41467-021-21220-5

<sup>1</sup>Yale School of Medicine, Yale University, New Haven, CT, USA;  
<sup>2</sup>Center for Science in the Public Interest, Washington, DC, USA;  
<sup>3</sup>Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA; <sup>4</sup>Harvey Cushing/John Hay Whitney Medical Library, Yale University, New Haven, CT, USA; <sup>5</sup>Section of General Medicine and the National Clinician Scholars Program, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA; <sup>6</sup>Center for Outcomes Research and Evaluation, Yale-New Haven Health System, New Haven, CT, USA; <sup>7</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA; <sup>8</sup>Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA, joshua.wallach@yale.edu

**Conflicts of Interest Disclosures** Joseph J. Ross is the US Outreach and Associate Research Editor at *The BMJ* and currently

receives research support through Yale University from Johnson & Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology, from the US Food and Drug Administration for the Yale–Mayo Clinic Center for Excellence in Regulatory Science and Innovation program (grant U01FD005938), from the Agency for Healthcare Research and Quality (grant R01HS022882), from the National Heart, Lung, and Blood Institute of the National Institutes of Health (grants R01HS025164 and R01HL144644), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International; in addition, Joseph J. Ross is an expert witness at the request of the relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. Joshua D. Wallach currently receives research support from the US Food and Drug Administration, the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award K01AA028258, and through Yale University from Johnson & Johnson to develop methods of clinical trial data sharing. No other disclosures were reported.

**Funding/Support** Osman Moneer received support from the US Food and Drug Administration through the Yale–Mayo Center for Excellence in Regulatory Science and Innovation Scholars Program.

**Role of the Funder/Sponsor** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Peer Review in a General Medical Research Journal Before and During the COVID-19 Pandemic

Roy H. Perlis,<sup>1,2</sup> Jacob Kendall-Taylor,<sup>3</sup> Ishani Ganguli,<sup>2,4</sup> Kamber Hart,<sup>5</sup> Jesse A. Berlin,<sup>2,6</sup> Steven M. Bradley,<sup>2,7</sup> Sébastien Haneuse,<sup>2,8</sup> Sharon K. Inouye,<sup>2,9</sup> Elizabeth A. Jacobs,<sup>2,10</sup> Arden Morris,<sup>2,11</sup> Eli Perencevich,<sup>2,12</sup> Lawrence N. Shulman,<sup>2,13</sup> N. Seth Trueger,<sup>2,14</sup> Stephan D. Fihn,<sup>2,15</sup> Frederick P. Rivara,<sup>2,15</sup> Annette Flanagin<sup>3</sup>

**Objective** Although peer review is an essential component of publication for new research, the viability of this process has been questioned, particularly with the added stressors of the COVID-19 pandemic. This study characterized rates of peer reviewer acceptance of invitations to review manuscripts, reviewer turnaround times, and editor-assessed quality of reviews before and after the start of the COVID-19 pandemic at a large, open-access general medical journal.

**Design** This retrospective, pre-post cohort study examined all research manuscripts submitted to *JAMA Network Open* between January 1, 2019, and June 29, 2021, either directly or via transfer from other *JAMA Network* journals, for which at least 1 peer review of manuscript content was solicited. Measures were compared between the period prior to the World Health Organization declaration of a COVID-19 pandemic on March 11, 2020 (14.3 months), and the period during the pandemic (15.6 months) among all reviewed manuscripts, and between pandemic-period manuscripts that did or did not address COVID-19. For each reviewed manuscript, the number of invitations sent to reviewers, proportions of reviewers accepting invitations, time in days to return reviews, and editor-assessed quality ratings of reviews were determined.

**Results** In total, the journal sought review for 5013 manuscripts, including 4295 Original Investigations (85.7%) and 718 Research Letters (14.3%); 1860 manuscripts were submitted during the prepandemic period and 3152 during the pandemic period. Overall mean (SD) volume of manuscripts reviewed per week increased from 30.3 (8.6) to 46.4 (12.2) manuscripts;  $P < .001$ . Comparing the prepandemic period with the pandemic period, a greater proportion of invited reviewers declined to review during the pandemic (from 33.0% to 34.5%;  $P = .02$ ), and the mean (SD) number of reviewer invitations per manuscript increased (from 5.99 [3.57] to 6.99 [4.46];  $P < .001$ ). However, the mean (SD) number of reviews rated as high quality (very good or excellent) per manuscript increased from 1.28 (0.72) to 1.48 (0.68), and the mean (SD) time for reviewers to return reviews was modestly shorter (from 15.82 [7.61] days to 14.35 [6.95] days;  $P < .001$ ), a difference that persisted in regression models accounting for manuscript type and topic.

**Conclusions** In this cohort study, peer reviewers were less likely to accept invitations to review manuscripts during the pandemic, but the speed and editor-reported quality of reviews improved. Additional study encompassing a broader set of journals will be necessary to understand the generalizability of these results, and to clarify how the pandemic has affected reviewer burden and fatigue.

<sup>1</sup>Harvard Medical School and Massachusetts General Hospital, Boston, MA, USA, rperlis@mgh.harvard.edu; <sup>2</sup>*JAMA Network Open*, Chicago, IL, USA; <sup>3</sup>*JAMA Network*, Chicago, IL, USA; <sup>4</sup>Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA; <sup>5</sup>Harvard Medical School, Boston, MA, USA; <sup>6</sup>Rutgers School of Public Health, Piscataway, NJ, USA; <sup>7</sup>Minneapolis Heart Institute and Minneapolis Heart Institute Foundation, Minneapolis, MN, USA; <sup>8</sup>Harvard T. H. Chan School of Public Health, Boston, MA, USA; <sup>9</sup>Harvard Medical School, Hebrew

SeniorLife, and Beth Israel Deaconess Medical Center, Boston, MA, USA; <sup>10</sup>MaineHealth and Maine Medical Center Research Institute, Scarborough, ME, USA; <sup>11</sup>Stanford University School of Medicine, Stanford, CA, USA; <sup>12</sup>Carver College of Medicine, University of Iowa, Iowa City, IA, USA; <sup>13</sup>Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA; <sup>14</sup>Department of Emergency Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; <sup>15</sup>University of Washington School of Medicine, Seattle, WA, USA

**Conflict of Interest Disclosures** Roy H. Perlis, Ishani Ganguli, Jesse A. Berlin, Steven M. Bradley, Sebastien Haneuse, Sharon K. Inouye, Elizabeth A. Jacobs, Arden Morris, Eli Perencevich, Lawrence N. Schulman, N. Seth Trueger, Stephan D. Fihn, and Frederick P. Rivara report receiving financial support from *JAMA Network* for service as editors at *JAMA Network Open*. Jacob Kendall-Taylor and Annette Flanagan are paid editorial staff for the *JAMA Network*. Annette Flanagan is executive director of the Peer Review Congress but was not involved in the review or decision of this abstract. Roy M. Perlis additionally reports receiving personal fees for service on scientific advisory boards for Belle Artificial Intelligence, Burrage Capital, Circular Genomics, Genomind, Psy Therapeutics, and RID Ventures and holding equity in Belle Artificial Intelligence, Psy Therapeutics, and Circular Genomics.

## Assessing the Readability and Quality of Patient or Caregiver Fact Sheets for COVID-19 Therapeutics with Emergency Use Authorization by the Food and Drug Administration

Shelly Melissa Pranić,<sup>1</sup> Jasna Karacić<sup>2</sup>

**Objective** Aside from remdesivir, the first US Food and Drug Administration (FDA)—approved drug to treat COVID-19, the FDA has authorized on an emergency basis (known as an emergency use authorization [EUA]) the use of drugs given the relative unavailability of effective COVID-19 treatments.<sup>1</sup> This study was conducted to determine fact sheet readability and quality of FDA-approved and EUA nonvaccine drugs or biologicals (therapeutics) to treat COVID-19.

**Design** In a cross-sectional study, facts sheets with 1 or more issuances using Google and the term *fact sheet* and the therapeutic name and FDA EUA website for fact sheets for patients, parents, or caregivers on January 17, 2022, and March 2, 2022, were identified.<sup>1</sup> Similarities in quality and readability between fact sheets allowed grouping by therapeutic. Two investigators independently selected eligible English-language fact sheets on FDA-approved drugs and EUAs for COVID-19 treatment. Primary outcomes were readability and quality. Seven readability tests were used including (1) Flesch-Kincaid reading ease (FKRE) index (ranging from 0 to 100 with higher scores corresponding to reading ease); (2) Flesch-Kincaid grade (FKG) level (ranging from grades 0 to 18 [college graduate], with lower grades corresponding to easier readability); (3) Gunning-Fog (GF) score (ranging from grades 0 to 20 [college graduate]); (4) Coleman-Liau index (CLI; ranging from grade 4 to college graduate); (5) automated readability index (ARI; ranging from grades 5 to 22 [college graduate]); (6) New Dale-Chall Readability (NDCR; ranging from grade 4 to college graduate); and (7) simple measure of gobbledegook (SMOG) index (ranging from grade 3 to college graduate). Secondary outcomes were word, syllable, and sentence counts.

**Table 43. Readability Scores and Other Characteristics of Patient, Parent, or Caregiver Fact Sheets for COVID-19 Therapeutics With US Food and Drug Administration Emergency Use Authorization According to Therapeutic Type**

Readability score	Median (range)			
	Total (N = 18)	Type of COVID-19 therapeutics		
		Antiviral (n = 6) <sup>a</sup>	SARS-CoV-2 monoclonal antibody (n = 9) <sup>b</sup>	Immune modulator (n = 3) <sup>c</sup>
Flesch-Kincaid reading ease index	47.80 (38.59-54.18)	48.81 (46.35-52.21)	44.92 (38.59-49.03)	49.43 (43.11-54.18)
Flesch-Kincaid grade level	10.33 (7.81-12.15)	10.17 (9.40-11.14)	11.07 (9.35-12.15)	10.57 (7.81-11.59)
Gunning-Fog score	12.08 (9.42-14.00)	11.67 (10.96-12.56)	12.32 (11.34-13.99)	12.04 (9.42-12.68)
Coleman-Liau index	12.00 (10.26-12.86)	11.57 (11.16-12.06)	12.00 (11.66-12.86)	11.39 (10.26-12.74)
Automated readability index	10.00 (6.48-11.65)	9.60 (8.57-11.21)	9.96 (9.62-11.58)	10.28 (6.48-11.65)
Dale-Chall Readability score	6.43 (5.69-7.12)	6.05 (5.88-6.51)	6.68 (6.35-7.03)	6.25 (5.69-7.12)
Simple measure of gobbledegook index	12.82 (10.21-14.13)	12.61 (12.38-13.64)	13.02 (12.51-14.13)	13.24 (10.21-13.34)
Word count	1402.0 (1083.0-1853.0)	1316.00 (1168-1758)	1463.00 (1110-1853)	1250.00 (1083-1776)
Syllables per word	1.7 (1.6-1.8)	1.7 (1.7-1.7)	1.7 (1.7-1.8)	1.7 (1.7-1.7)
Words with >4 syllables	25.50 (11-93)	19.00 (11-84)	27.00 (13-93)	29.00 (13-46)
Words per sentence	16.10 (8.50-18.80)	14.45 (13.20-18.40)	16.80 (14.20-18.80)	17.10 (8.50-17.80)
Sentence count	91.50 (62-147)	82.00 (79-134)	93.00 (78-105)	105.00 (62-147)
DISCERN score <sup>d</sup>	48.50 (46-57)	53.50 (47-57)	48.00 (46-54)	48 (46-52)

<sup>a</sup>Remdesivir, molnupiravir, and nirmatrelvir/ritonavir.

<sup>b</sup>Bamlanivimab/etesevimab, casirivimab/imdevimab, sotrovimab, tixagevimab/cilgavimab, and bebtelovimab.

<sup>c</sup>Tocilizumab, baricitinib, and convalescent plasma.

<sup>d</sup>The quality of the information was classified according to the median score as excellent (63-80), good (51-62), fair (39-50), poor (28-38), or very poor (<27).

Agreement between investigators was good (80%) in rating 2 fact sheets for quality using the 16-item DISCERN instrument<sup>2</sup> with Likert responses (1 indicates minimum and 5, maximum) where the total is 80 and lowest is 16, corresponding to low-quality information; another investigator rated the remainder. Items on the DISCERN instrument assess the transparency of authorship information and relevance of treatment options to patients as described previously.<sup>2</sup>

**Results** Overall, 18 fact sheets were found that described 6 antiviral (37.5%) (4 for remdesivir and 1 each molnupiravir and nirmatrelvir/ritonavir), 9 SARS-CoV-2-targeting monoclonal antibody (43.8%) (1 for bamlanivimab/etesevimab, 3 for casirivimab/imdevimab, 2 each for sotrovimab, tixagevimab/cilgavimab, and 1 for bebtelovimab), and 3 immune modulator (21.4%) (1 each for tocilizumab, baricitinib, and convalescent plasma) information. **Table 43** shows the median FKRE, FKG, GF, CLI, ARI, NDCR, and SMOG reading levels above the sixth grade; quality was fair.

**Conclusions** Although of fair quality, the reading grade level of fact sheets intended for patients, parents, or caregivers for COVID-19 therapeutics was high, reflecting a need for FDA officials to enforce readable resources from drug manufacturers.

## References

1. US Food and Drug Administration. Emergency use authorization: drugs and non-vaccine biological products. Accessed March 2, 2022. <https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization#coviddrugs>
2. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health*. 1999;53(2):105-111. doi:10.1136/jech.53.2.105

<sup>1</sup>University of Split School of Medicine, Cochrane Croatia, Split, Croatia, shelly.pranic@mefst.hr; <sup>2</sup>Croatian Association for the Promotion of Patients' Rights, Cochrane Croatia, Split, Croatia

**Conflict of Interest Disclosures** None reported.

## Examination of Adapting the Patient-Centered Outcomes Research Institute's Multistakeholder Application Review Processes During COVID-19

Laura P. Forsythe,<sup>1</sup> Robin Bloodworth,<sup>1,2</sup> Carolyn Mohan,<sup>1</sup> Rachel C. Hemphill,<sup>1</sup> Esther Nolton,<sup>1</sup> Ponta Abadi,<sup>1,3</sup> Lisa Stewart,<sup>1,4</sup> Krista Woodward<sup>1,5</sup>

**Objective** The Patient-Centered Outcomes Research Institute's (PCORI's) review of research applications uniquely includes patient and stakeholder reviewers alongside scientists. PCORI switched to virtual panel discussions in response to the COVID-19 pandemic. The few prior studies examining virtual review for health research were mostly small scale, provided mixed results, and did not consider multistakeholder processes.<sup>1,2</sup> This study examined how virtual panels compared with in-person panels on reviewer scores and experiences, whether differences between panels varied by reviewer type (scientist, patient, or stakeholder), and reviewer perceptions of challenges and benefits of virtual panels.

**Design** This cross-sectional, mixed-methods study analyzed data for PCORI funding opportunities before and after switching to virtual review, including review score data (8 in-person cycles and 4 virtual, 2017-2021) and closed and open-ended responses from anonymous online surveys of reviewers (1 in-person cycle [2017] and 2 virtual [2020-2021]). Virtual vs in-person panels were compared on (1) final overall review scores and changes in overall scores before and after panel discussion for primary reviewers using linear regression, which included examining the interaction of panel type and reviewer type; and (2) reviewer perspectives on giving and receiving input using logistic regression (5-point Likert agreement scales dichotomized as agree vs neutral/disagree). Regression models controlled for reviewer and

**Table 44. Regression Models Comparing Virtual vs In-Person Discussion Panels for PCORI Merit Review**

	Adjusted β (SE)		Adjusted odds ratio (95% CI)	
	Overall scores <sup>a</sup>	Change in scores <sup>b</sup>	Agreement that “overall, patient/stakeholder reviewers were receptive to input from scientific reviewers” <sup>c</sup>	Agreement that “overall, scientific reviewers were receptive to input from patient/stakeholder reviewers” <sup>c,d</sup>
<b>Panel type</b>				
In-person	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Virtual	-0.061 (0.06)	-0.033 (0.03)	0.585 (0.202-1.694)	0.827 (0.362-1.888)
<b>Reviewer type</b>				
Scientist	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Patient	0.095 (0.05)	0.146 (0.03) <sup>e</sup>	1.140 (0.362-3.590)	0.367 (0.131-1.070)
Stakeholder	0.135 (0.05) <sup>g</sup>	0.116 (0.03) <sup>e</sup>	4.601 (0.593,35.70) <sup>f</sup>	0.276 (0.104-0.732) <sup>g</sup>
<b>Interaction</b>				
In-person by scientist	1 [Reference]	1 [Reference]	NA	NA
In-person by patient	1 [Reference]	1 [Reference]	NA	NA
In-person by stakeholder	1 [Reference]	1 [Reference]	NA	NA
Virtual by scientist	1 [Reference]	1 [Reference]	NA	NA
Virtual by patient	0.036 (0.12)	0.046 (0.07)	NA	NA
Virtual by stakeholder	0.091 (0.12)	0.081 (0.07)	NA	NA

Abbreviations: NA, not applicable; PCORI, Patient-Centered Outcomes Research Institute; PFA, PCORI funding announcement.

<sup>a</sup>Adjusted for PFA type, repeat reviewer status, resubmission status, and preliminary score.

<sup>b</sup>Adjusted for PFA type, repeat reviewer status, resubmission status, preliminary score, and score change direction (change toward weaker score, no change, or change toward stronger score).

<sup>c</sup>Adjusted for PFA type and repeat reviewer status; models for survey items did not include interaction terms owing to limited sample sizes.

<sup>d</sup>Adjusted for PFA type; this model was not adjusted for repeat reviewer status because smaller sample sizes and skewness of data did not allow for the calculation of coefficients and estimates.

<sup>e</sup>Significant at  $P < .001$ .

<sup>f</sup>Few stakeholder reviewers disagreed with this survey item, which resulted in imprecise coefficients and estimates.

<sup>g</sup>Significant at  $P = .01$ .

application characteristics. Sensitivity analyses included multilevel models to account for hierarchy of reviewers and scores. For virtual cycles, open-ended survey responses were analyzed through inductive and deductive qualitative coding to identify themes regarding reviewers' perceived challenges and benefits of virtual review.

**Results** The analytic sample included 2897 reviews (2253 in-person, 644 virtual) and 388 survey responses (191 in-person, 197 virtual; 75%-83% response rate). Final review scores (mean [SD] score for in-person, 4.7 [1.67]; for virtual, 4.5 [1.67]) and absolute value of score changes (mean [SD] score for in-person, 0.8 [0.96]; for virtual, 0.7 [0.97]) were similar between virtual and in-person panels ( $P > .05$  for all) (**Table 44**); there were no significant associations for interactions of panel type and reviewer type ( $P > .05$ ). In closed-ended survey items, most reviewers agreed that reviewers of each type (patient or stakeholder and scientist) were receptive to input from the other type (85%-96% across reviewer types), and there were no differences in agreement by panel type ( $P > .05$  for all) (**Table 44**). In open-ended survey responses, reviewers noted challenges of virtual panels, including disruptions to discussion quality and flow, missing social interactions among reviewers, and technical and logistical issues; reviewers also noted benefits of virtual panels, including convenience and lack of travel.

**Conclusions** Findings indicate that, despite some challenges, virtual review panels were similar to in-person

panels on review scores and key aspects of reviewer experiences in a multistakeholder process. Virtual panels could be further considered as a viable approach for PCORI in the future to offer flexibility in circumstances beyond the COVID-19 pandemic.

## References

1. Carpenter AS, Sullivan JH, Deshmukh A, Glisson SR, Gallo S. A retrospective analysis of the effect of discussion in teleconference and face-to-face scientific peer-review panels. *BMJ Open*. 2015;5:e009138. doi:10.1136/bmjopen-2015-009138
2. Gallo SA, Carpenter AS, Glisson SR. Teleconference versus face-to-face scientific peer review of grant application: effects on review outcomes. *PLoS ONE*. 2013;8(8):e71693. doi:10.1371/journal.pone.0071693

<sup>a</sup>Patient-Centered Outcomes Research Institute (PCORI), Washington, DC, USA, lforsythe@pcori.org; <sup>b</sup>Abt Associates, Rockville, MD, USA; <sup>c</sup>OCHIN Inc, Portland, OR, USA; <sup>d</sup>Discovery USA, Chicago, IL, USA; <sup>e</sup>Department of Population, Family, and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** PCORI provided funding for this work.

**Role of the Funder/Sponsor** PCORI staff planned and conducted all of the following aspects of this work: design and conduct of the study; collection, management, analysis, and

interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgments** PCORI thanks everyone who has served as a Merit Reviewer for their time and invaluable input on research applications. The authors also thank Cary Scheiderer and Layla Lavasani (former PCORI staff) for sharing their insights about key elements of the review process and Sarah Cohen and Heidi Reichert of EpidStrategies for their consultation on statistical methods used in this abstract.

## Peer Review

### Association of Peer Review With Completeness of Reporting, Transparency for Risk of Bias, and Spin in Diagnostic Test Accuracy Studies Published in Imaging Journals

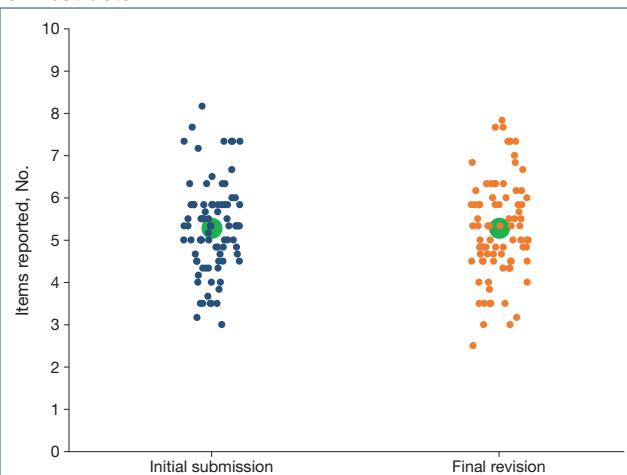
Sakib Kazi,<sup>1</sup> Robert A. Frank,<sup>2</sup> Jean-Paul Salameh,<sup>3,4</sup> Nicholas Fabiano,<sup>1</sup> Marissa Absi,<sup>1</sup> Alex Pozdnyakov,<sup>5</sup> Nayaar Islam,<sup>4,6</sup> Daniël A. Korevaar,<sup>7</sup> Jérémie F. Cohen,<sup>8</sup> Patrick M. Bossuyt,<sup>9</sup> Mariska M. G. Leeflang,<sup>9</sup> Kelly D. Cobey,<sup>10</sup> David Moher,<sup>10</sup> Mark Schweitzer,<sup>11</sup> Yves Menu,<sup>12</sup> Michael Patlas,<sup>5</sup> Matthew D. F. McInnes<sup>4,5</sup>

**Objective** To evaluate whether peer review of diagnostic test accuracy (DTA) studies published by imaging journals is associated with changes in completeness of reporting, transparency of risk of bias, and spin, given that there is limited evidence to support the concept that peer review improves the completeness of research reporting.<sup>1,2</sup>

**Design** This retrospective cross-sectional study evaluated articles published in the *Journal of Magnetic Resonance Imaging* (*JMRI*; 2019 impact factor [IF], 4.0), the *Canadian Association of Radiologists Journal* (*CARJ*; IF, 1.7), and *European Radiology* (*EuRad*; IF, 4.1) before March 31, 2020.<sup>3</sup> Initial submitted and final versions of manuscripts were screened consecutively in reverse chronological order to include a minimum of 23 articles (based on power calculation) per journal. At least 30 eligible articles from each journal were collected when available to account for potential exclusions. Primary studies evaluating the diagnostic accuracy of an imaging test in humans were included. Studies exclusively reporting on prognostic or predictive tests were excluded. Studies were evaluated independently by 2 reviewers blinded to version for completeness of reporting using the Standards for Reporting Diagnostic Accuracy Studies (STARD) 2015 and STARD for Abstracts guidelines, transparency of reporting for risk of bias assessment based on the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2), and actual and potential spin using modified published criteria. Two-tailed paired *t*-tests and paired Wilcoxon signed-rank tests were used for comparisons; *P* < .05 was considered statistically significant.

**Results** Of 692 diagnostic accuracy studies screened, 84 articles published in 2014 to 2020 from 3 journals were included: *JMRI*, 30 articles; *CARJ*, 23; and *EuRad*, 31. Reporting by STARD 2015 increased between initial submissions and final accepted versions (mean reported

**Figure 20. Completeness of Reporting Using the Standards for Reporting Diagnostic Accuracy Studies 2015 Guidelines for Abstracts**



items 16.67 vs 17.47; change, 0.80 [95% CI, 0.25 to 1.17]; *P* = .002). From STARD, *sources of funding and other support* (item 30.1) and *role of funders* (item 30.2) had the largest change of 0.32 (*P* < .001). No difference was found for the reporting of STARD for Abstracts (5.28 vs 5.25; change, -0.03; 95% CI, -0.15 to 0.11; *P* = .74); QUADAS-2 (6.08 vs 6.11; 0.03; 95% CI, -1.00 to 0.50; *P* = .92); actual spin (2.36 vs 2.40; change, 0.04; 95% CI, 0.00 to 1.00; *P* = .39); or potential spin practices (2.93 vs 2.81; change, -0.12; 95% CI, -1.00 to 0.00; *P* = .23) (Figure 20).

**Conclusions** This retrospective cross-sectional study found that peer review was associated with a marginal improvement in completeness of full text; however, it was not associated with abstract reporting in published imaging DTA studies nor with improvement in transparency for risk of bias assessment or reduction in spin. Considering that this study included articles from only 3 radiology journals, the findings may not be generalizable to other journals, other fields of DTA research, or non-DTA study designs. Interventions such as reviewer training and use of checklists should be evaluated.

## References

- Jefferson T, Rudin M, Brodney Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. Cochrane Library. 2020. doi:10.1002/14651858.MR000016.pub3
- Bruce R, Chauvin A, Trinquart L, Ravaud P, Boutron I. Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. *BMC Medicine*. 2016;14(1):1-16. doi:10.1186/s12916-016-0631-5
- Clarivate Analytics. Journal Citation Reports. 2019 Journal Impact Factor. Accessed July 11, 2020. <https://clarivate.com/blog/announcing-the-2019-journal-citation-reports/>

<sup>1</sup>Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada; <sup>2</sup>Department of Radiology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada; <sup>3</sup>Faculty of Health Sciences, Queen's University, Ottawa, Ontario, Canada; <sup>4</sup>Clinical

Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, mmcinnies@toh.ca; <sup>5</sup>Department of Radiology, McMaster University, Hamilton, Ontario, Canada; <sup>6</sup>School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada; <sup>7</sup>Department of Respiratory Medicine, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, the Netherlands; <sup>8</sup>Department of Pediatrics Inserm UMR 1153, Centre of Research in Epidemiology and Statistics, Necker—Enfants Malades Hospital, Assistance Publique—Hôpitaux de Paris, Université de Paris, Paris, France; <sup>9</sup>Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands; <sup>10</sup>Department of Radiology, Wayne State University School of Medicine, Detroit, MI, USA; <sup>12</sup>Department of Radiology, Sorbonne Université—AP-HP, Paris, France

**Conflict of Interest Disclosures** Mark Schweitzer, Yves Menu, Michael Patlas, and Kelly D. Cobey have active affiliations with the 3 journals used as data sources but had no role in data extraction, analysis, or interpretation, but reviewed and approved the work. Michael Patlas reported an editorial honorarium from Springer outside of the submitted work. No other disclosures were reported.

**Funding/Support** Funding support was received from the Philips Radiological Society of North America research seed grant (RSNA Research & Education Foundation), Mitacs Research Training Award, and the Department of Radiology MD Summer Student Fund at the University of Ottawa. Study performance and manuscript content were the sole task and responsibility of the investigators and do not necessarily represent the official views of the funders.

**Role of the Funder/Sponsor** The funders had no role in data collection, analysis, interpretation, or manuscript composition.

**Acknowledgments** Sakib Kazi and Robert A. Frank contributed equally to this work.

### A Survey of Reviewers' Perspectives on Options for Open and Transparent Peer Review at *Annals of Internal Medicine*

Jill Jackson,<sup>1</sup> Christine Laine,<sup>1</sup> Julie Kostelnik<sup>1</sup>

**Objective** In open transparent peer review, the authors and reviewers know each other's identities, and published articles are accompanied by the signed reviewer comments. *Annals of Internal Medicine* (*Annals*) uses single-blind review—reviewers know authors' identities, but reviewers' identities are not disclosed to authors and reviews are not published. The objective of this survey was to examine whether moving to more open peer review could affect the willingness of *Annals* peer reviewers to review for the journal or the nature of their comments.

**Table 45. Willingness of Respondents to Continue to Review With an Open Review Model**

	Respondents who were somewhat or very unlikely to review, %	Respondents who were indifferent, %	Respondents who were somewhat or very likely to review, %
Reviewer identity disclosed to authors	35	13	52
Reviewer identity published with article	28	16	56
Review published with article	28	21	51

**Design** The authors conducted an online survey of 5977 persons who reviewed for the journal from January 1, 2019, to January 1, 2022. In addition to information on age, specialty, and professional setting and experience, the survey asked how likely (on a 5-point scale) respondents would be to continue to review for *Annals* if they were asked to sign their comments for authors, if the published article identified the reviewers, and if the reviews accompanied published articles. In addition, the survey asked whether and how open review would influence the nature of their comments.

**Results** Of 1421 respondents (24% response rate), 71% were more than 10 years past completing training, 48% reported being researchers or clinician researchers, 61% peer reviewed at least 5 times in a typical year, and 66% had authored more than 20 peer-reviewed publications. Approximately one-third reported that they would be unlikely to review if *Annals* adopted an open review policy (**Table 45**). Whereas 42% reported that this model would not affect their comments, 10% reported that their reviews would be less detailed and 28% reported that their reviews would be less critical. However, 17% reported that their comments would be more detailed, 3% would be more critical of the article, and 20% responded with personal answers to the question.

**Conclusions** An open review model could adversely affect the willingness of current *Annals* peer reviewers to continue to review and could alter the nature of reviewer comments.

<sup>1</sup>*Annals of Internal Medicine*, American College of Physicians, Philadelphia, PA, USA, jjackson@acponline.org

**Conflict of Interest Disclosures** Christine Laine is a member of the Peer Review Congress advisory board but was not involved in the review or decision of this abstract. No other disclosures reported.

### Association Between Peer Reviewers' Priority Ratings of Impact of Research Manuscripts With Citations and Altmetric Scores of Subsequently Published Articles in the *Journal of Medical Internet Research*

Gunther Eysenbach<sup>1</sup>

**Objective** Peer-reviewed journals ask reviewers to rate the perceived impact or priority of a manuscript. Previous research has suggested an association between reviewer priority scores and citations.<sup>1</sup> Altmetrics (alternative metrics) provide an alternative view on social impact (ie, uptake on social media, in policy documents, or by news articles<sup>2</sup>); however, their association with reviewer scores has not been explored.<sup>3</sup> It is unclear whether reviewer ratings more closely reflect scientific impact (ie, citations) or social impact (altmetric scores), and which metric is more valid to reflect reviewers' priority ratings.

**Design** This was a longitudinal bibliometric cohort study that followed 451 original research articles published in 2018 in the *Journal of Medical Internet Research*, a health services research and medical informatics journal (impact factor, >5). This journal asks peer reviewers to rate the priority (defined

as potential impact) of a manuscript on an ordinal rating scale with possible scores of 1, 2, 5, and 10 (highest priority). Manuscripts are typically reviewed by 2 reviewers. The mean priority score of all reviewers for a manuscript in the first review round constitutes the Manuscript Average Priority Score (MAPS). For this analysis, manuscripts were categorized into 4 quartiles (Qs), with the groups labeled as Q4 (MAPS score,  $\leq 3$ ) to Q1 (MAPS score,  $> 5$ ). The dependent variables, citations, and altmetric scores were obtained from the Dimensions database in February 2022; manuscripts and published articles were similarly stratified into quartiles, with the citation (or altmetrics) quartile Q1 containing the group of articles with the highest citation count (or altmetric score). The association between independent variables (MAPS scores) and citation or altmetric scores was measured using  $\chi^2$  tests for  $4 \times 4$  contingency tables for the quartiles and using Spearman rank correlation between MAPS score ranks and citation or altmetric rank, respectively.

**Results** The MAPS scores for 451 published articles ranged from 1.5 to 10; citations, from 0 to 253; and altmetric scores, from 1 to 849. Although both mean and median citations as well as altmetric scores were higher in the higher MAPS quartiles (**Table 46**), the results of  $\chi^2$  tests were not statistically significant for citations ( $P = .46$ ) but were statistically significant for altmetric scores ( $P = .03$ ). The Spearman rank correlation between citation ranks and MAPS score ranks was statistically significant but weak ( $\rho = .0955$ ;  $r^2 = .009$ ;  $P = .03$ ). In contrast, altmetric score ranks had a stronger correlation with MAPS score ranks ( $\rho = .1313$ ;  $r^2 = .017$ ;  $P = .002$ ).

**Conclusions** This longitudinal bibliometric cohort study found that in the *Journal of Medical Internet Research*, a journal whose subject matter lends itself to the type of attention measured by altmetrics, altmetric scores seemed to be better correlated than citations with a manuscript's potential impact as assessed by reviewers. Peer reviewers may interpret priority and impact in terms of social impact, rather than citations, raising further questions about the

**Table 46. Descriptive Statistics of Citations and Altmetrics by Reviewer Priority (MAPS) Score Category**

Reviewer priority category (MAPS range)	No. (N = 451)	Mean (SD)	Median (IQR)
Citations, No.			
Q4 (1.5 to 3)	117	14.5 (11.6)	11.0 (7.0-19,012.0)
Q3 (>3 to $\leq 3.5$ )	109	23.0 (22.4)	18.0 (9.0-28,019.0)
Q2 (>3.5 to $\leq 5$ )	154	33.2 (26.9)	25.0 (15.3-40,825.5)
Q1 (>5)	71	52.3 (49.6)	35.0 (19.0-65,546.5)
Altmetric score			
Q4 (1.5 to 3)	117	3.9 (1.5)	4.0 (3.0-5.02.0)
Q3 (>3 to $\leq 3.5$ )	109	8.2 (2.7)	8.0 (7.0-1003.0)
Q2 (>3.5 to $\leq 5$ )	154	34.2 (51.4)	21.0 (16.0-26,010.0)
Q1 (>5)	71	88.6 (125.2)	49.0 (34.0-83,049.0)

Abbreviations: MAPS, Manuscript Average Priority Score (scale: 1, 2, 5, 10 [highest impact/priority]; Q, quartile).

appropriateness of citation-based metrics to measure impact as understood by reviewers.

## References

- Ophof T, Coronel R, Janse MJ. The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovasc Res*. 2002;56(3):339-346. doi:10.1016/S0008-6363(02)00712-5
- Eysenbach G. Can tweets predict citations? metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res*. 2011;13(4):e123. doi:10.2196/jmir.2012
- Araujo AC, Vanin AA, Nascimento DP, et al. What are the variables associated with altmetric scores? *Syst Rev*. 2021;10:193. doi:10.1186/s13643-021-01735-0

<sup>1</sup>JMIR Publications, Toronto, ON, Canada, geysenba@gmail.com

**Conflict of Interest Disclosures** Gunther Eysenbach reported equity in JMIR Publications.

## Peer Review Process and Models

### Feasibility of a Peer Review Intervention to Reduce Undisclosed Discrepancies Between Registrations and Publications

TARG Meta-Research Group & Collaborators

Robert T. Thibault,<sup>1,2,3</sup> Tom E. Hardwicke,<sup>4</sup> Robbie W. A. Clark,<sup>2,3</sup> Charlotte R. Pennington,<sup>5,6</sup> Gustav Nilssonne,<sup>7,8</sup> Aoife O'Mahony,<sup>9</sup> Katie Drax,<sup>2,3</sup> Jacqueline Thompson,<sup>2,3</sup> Marcus R. Munafò<sup>2,3</sup>

**Objective** The authors developed a peer review intervention to reduce undisclosed discrepancies between study registrations and their associated publications, which are common.<sup>1,2</sup> The aim of this study was to (1a) evaluate the feasibility of incorporating discrepancy review as a regular practice at scientific journals; (1b) evaluate the feasibility of conducting a trial on discrepancy review; (2) explore the benefits and time required to incorporate discrepancy review as a regular practice at scientific journals; and (3) refine the discrepancy review process.

**Design** The authors invited editors in chief of 18 journals in medicine or psychology to participate and provided volunteer early-career researchers to act as peer reviewers who were specifically assigned to check for undisclosed discrepancies between registrations and submitted manuscripts of any study design. The authors called this process *discrepancy review*.

**Results** Of the 18 invited journals, 5 agreed to participate, 2 of which did not receive any manuscripts reporting a registration during the study period and 1 of which had difficulty adding discrepancy review to their manuscript handling procedures and therefore did not provide any manuscripts to review, leaving 2 participating journals.

**Table 47. Characteristics of the Manuscripts Reviewed**

	Clinical trial registries <sup>a</sup>	OSF	PROSPERO
Total manuscripts reviewed, No.	12	7	2
Original discrepancy review			
Submitted to editor, No.	10	5	1
Second review, No. <sup>b</sup>	6	2	0
Time, median (range), min	105 (16-180)	210 (90-360)	90
Updated discrepancy review			
Submitted to editor, No.	2	2	1
Second review, No. <sup>b</sup>	6	5	2
Time, median (range), min	28 (10-60)	50 (20-92)	43 (12-50)
Nonpermanent registrations	0	4 <sup>c</sup>	0
Manuscripts correctly labeled as a secondary publication, No./total No.	0/5	0/0	0/0
Importance of addressing discrepancies, No./total No. of submitted manuscripts <sup>d</sup>			
Quite important	3/7	1/7	1/2
Somewhat important	4/7	3/7	1/2
Not important or no discrepancies	0/7	3/7	0/2

Abbreviations: OSF, Open Science Framework; PROSPERO, International Prospective Register of Systematic Reviews.

<sup>a</sup>ClinicalTrials.gov (n = 10), ANZCTR (Australian New Zealand Clinical Trials Registry) (n = 1), ISRCTN (International Standard Randomised Controlled Trial Number) (n = 1).

<sup>b</sup>A second team member performed discrepancy review for each manuscript to examine consistencies and differences among discrepancy reviewer reports.

<sup>c</sup>In addition, 2 submitted manuscripts had permanent registrations on the OSF Registries webpage (<https://osf.io/registries>) but only included a link to a nonpermanent version on an OSF home page (<https://osf.io>).

<sup>d</sup>Only assessed for manuscripts that were not secondary publications associated with a clinical trial registration.

Discrepancy review was performed between January 29 and May 18, 2021, on all registered studies submitted to *Nicotine and Tobacco Research* (n = 18) and on all registered studies for which the editor in chief of *European Journal of Personality* acted as action editor (n = 3). **Table 47** details the reviewed manuscripts and findings. Registrations were generally too imprecise to be effectively evaluated by the original discrepancy review process, which used a detailed and structured checklist. Thus, the authors developed an updated discrepancy review process that used a semi-structured format with 8 guiding questions regarding exploratory studies, proper registration, retrospective registration, hypotheses, independent variables, outcome measures, analyses, and additional discrepancies.

Discrepancy reviewers provided 59 comments on the 12 manuscripts that were accepted for publication. Authors fully addressed 31, partially addressed 10, and did not address 18. Optional questionnaires were completed by 5 of 13 action editors and 4 of 21 manuscript authors who showed no opposition to discrepancy review.

**Conclusions** It was feasible for 2 journals interested in discrepancy review to implement this process when provided with discrepancy reviewers. A full trial of discrepancy review would be needed to evaluate its effect on reducing undisclosed discrepancies, possibly stratified by clinical trial

vs Open Science Framework registration given differences in the detail required by each registry.

## References

- TARG Meta-Research Group and Collaborators. Estimating the prevalence of discrepancies between study registrations and publications: a systematic review and meta-analyses. *medRxiv*. Preprint posted online August 9, 2021. doi:10.1101/2021.07.07.21259868
- Goldacre B, Drysdale H, Dale A, et al. COMParE: a prospective cohort study correcting and monitoring 58 misreported trials in real time. *Trials*. 2019;20(118):1-16. doi:10.1186/s13063-019-3173-2

<sup>a</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford, CA, USA robert.thibault@stanford.edu; <sup>b</sup>School of Psychological Science, University of Bristol, Bristol, UK; <sup>c</sup>MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK; <sup>d</sup>Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; <sup>e</sup>School of Psychology, Aston University, Birmingham, UK; <sup>f</sup>Institute of Health and Neurodevelopment, Aston University, Birmingham, UK; <sup>g</sup>Department of Clinical Neuroscience, Karolinska Institutet, Solna, Sweden; <sup>h</sup>Department of Psychology, Stockholm University, Stockholm, Sweden; <sup>i</sup>School of Psychology, Cardiff University, Cardiff, UK

**Conflict of Interest Disclosures** Charlotte R. Pennington is the local network lead of the UK Reproducibility Network for Aston University. Gustav Nilssonne is a member of the Committee for Open Badges and served for several years as its chair. All other authors declare no conflict of interest.

**Funding/Support** Robert T. Thibault is supported by a general support grant awarded to METRICS from the Laura and John Arnold Foundation and postdoctoral fellowships from the Canadian Institutes of Health Research and the Fonds de recherche du Québec—Santé. Tom E. Hardwicke receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 841188. Katie Drax is supported by the John Climax Benevolent Fund. Robbie W. A. Clark is supported by a SWDTP ESRC PhD studentship. Robert T. Thibault, Robbie W. A. Clark, Katie Drax, Jacqueline Thompson, and Marcus R. Munafò are all part of the MRC Integrative Epidemiology Unit (MC\_UU\_00011/7).

**Role of the Funder/Sponsor** The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Experience With Select Crowd Review in Peer Review for *The Thoracic and Cardiovascular Surgeon*

Roman Gottardi,<sup>1</sup> Peter Henning,<sup>2</sup> Jessica Bogensberger,<sup>2</sup> Markus K. Heinemann<sup>3</sup>

**Objective** To evaluate the experience with a new peer review method,<sup>1</sup> Select Crowd Review (SCR), for *The Thoracic and Cardiovascular Surgeon*. Upon submission, authors are given a choice to accept or decline SCR. A “crowd” was created in part from the existing reviewer pool and in part newly recruited. If authors agree to SCR, anonymized PDFs of manuscripts are made accessible to the crowd upon invitation

per email via an online platform for 10 days. This intuitive platform enables the crowd to enter comments directly into the text; a formal structured review is not required. Reviewers give their comments anonymized. An SCR editor summarizes the annotations and gives a recommendation. Both the commented PDF and summary are sent back to the authors through the editor in chief. The aim is to achieve a rapid and broader, and thus fairer, review process.

**Design** All manuscript submissions from introduction in July 2021 until January 15, 2022, were analyzed regarding acceptance and quality. Anonymized cardiac manuscripts were sent to a preselected crowd of 45 reviewers and entered regular double-blinded peer review at the same time. Efficiency and performance of the crowd's reviews were compared with that of regular reviewers. For thoracic manuscripts, a crowd was not yet available during this pilot period.

**Results** Of 162 total submitted manuscripts, 72 (44.4%) were selected for SCR; 84 were cardiac manuscripts, 39 (46.4%) of which were selected for SCR. Ten of those had to be rejected without any review, and 29 finally entered SCR. A first review process was completed for 24 manuscripts. For 3 manuscripts, the crowd did not respond. In all remaining 21 papers, the crowd's recommendation concurred with that of the regular reviewers, leading to 8 rejections. Regular peer review took up to 5 weeks. Nine manuscripts underwent repeated SCR after revision. On average, 3 (range, 0-9) crowd members sent in reviews. In revisions, average response was worse, with mostly only 1 previous reviewer responding.

**Conclusions** SCR encountered good acceptance by authors. Because the first experience showed absolutely concordant recommendations within 10 days compared with the slower traditional review, thoracic manuscripts have been included to gain more experience. If positive feedback continues, SCR may become an established method of peer review in selected journals. It is certainly helpful to achieve a fast first evaluation. Efficiency apparently must be increased for re-review of revisions.

## Reference

1. List B. Crowd-based peer review can be good and fast. *Nature*. 2017;546(7656):9. doi:10.1038/546009a

<sup>1</sup>Mediclin Herzzentrum Lahr, Lahr, Germany; <sup>2</sup>Thieme Medical Publishers, Stuttgart, Germany; <sup>3</sup>Cardiovascular Surgery, Universitaetsmedizin Mainz, Mainz, Germany, heinemann@uni-mainz.de

**Conflict of Interest Disclosures** Peter Henning and Jessica Boggensberger are employees of Thieme Medical Publishers and thus have an interest to have this new peer review method, developed by Thieme, evaluated for further use. Roman Gottardi is the SCR editor and Markus Heinemann is the editor in chief of *The Thoracic and Cardiovascular Surgeon*. They have no conflicts.

---

## Comparison of Distributed Peer Review Enhanced by Machine Learning and Natural Language Processing and With Traditional Panel-Based Peer Review of Astronomy Proposals

Wolfgang Kerzendorf,<sup>1</sup> Ferdinando Patat,<sup>2</sup> Dominic Bordelon,<sup>3</sup> Glenn van de Ven,<sup>4</sup> Tyler Pritchard<sup>5</sup>

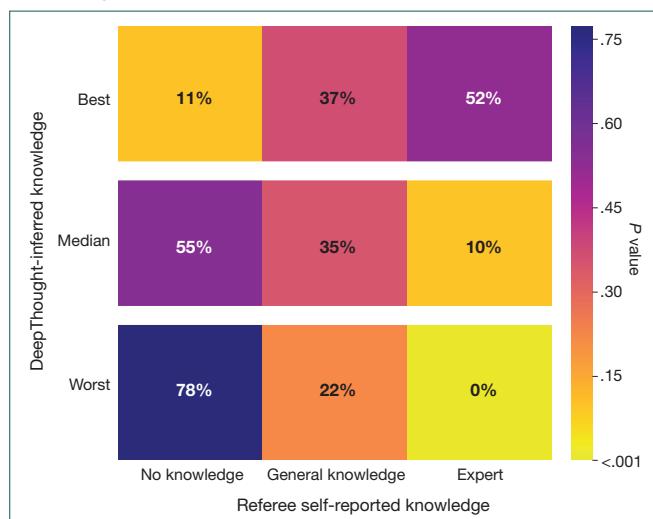
**Objective** An ever-rising number of researchers (which increased by 15% between 2014 and 2018<sup>1</sup>) has led to a substantial increase of publications and proposals in astronomy [that overwhelms traditional peer review systems]. Distributed peer review (DPR) is a model that uses the pool of proposers as the pool of reviewers and has been suggested<sup>2</sup> and used in astronomyX.<sup>3</sup> This was a quantitative study that explored a new language process-enhanced variation of the DPR scheme (named *DeepThoughtDPR*). This study tested whether a natural language, machine learning matching algorithm can identify knowledgeable reviewers and whether the DPR system has a similar interreviewer reliability compared with the traditional system.

**Design** This was a cross-sectional analysis involving astronomers who submitted telescope time proposals for European Southern Observatory Period 103 on September 27, 2018; in all, 172 volunteers submitted their grants specifically to the machine learning-enhanced DPR. Each proposal was reviewed by 8 peers, resulting in the same number of proposals needing to be reviewed by each proposer; this resulted in 1336 unique reviews (data set available at <https://zenodo.org/record/2634598>). Next, 112 randomly chosen volunteers from among the proposers reviewed proposals that were matched via an algorithm. The volunteers were given 4 “best match” proposals, 2 “median match” proposals, and 2 “worst match” proposals but were not told about this selection. The 112 volunteers were also asked to report their expertise on each proposal, and the data were used to test the matching algorithm. The other 60 volunteers used an essentially random selection scheme. The DPR scheme was compared with traditional panel-based review on the basis of interreviewer reliability by using 15,000 existing reviews from traditional panel-based review. The comparison tested in which quartile the proposal was ranked by independent groups of reviewers within the same review process.

**Results** Figure 21 shows the conditional probability calculated from 820 reviews. The volunteers reported they had “no knowledge” for the “worst match” proposals in 78% of cases, had “general knowledge” for the “median match” proposals in 35% of cases, and considered themselves “expert” for the “best match” proposals in 52% of cases. The match showed a Spearman rank coefficient of 0.64. The comparison of the interreviewer reliability between the traditional panel-based approach and the DPR approach showed a maximum difference of 5% when a quartile agreement matrix scheme was used.

**Conclusions** The reviewer-matching algorithm had a high probability of identifying cases of “no knowledge.” The lower probabilities of “expert” and “general knowledge” may be

**Figure 21. Accuracy of Self-reported vs DeepThought Knowledge**



The percentage in each square represents the conditional probability of the proposal rated as no knowledge, general knowledge, or expert given that DeepThought used the categories “best,” “median,” or “worst.” The conditional probability between self-reported and DeepThought-inferred knowledge had a Spearman rank coefficient of 0.64.

ascribed to self-efficacy but need to be tested in future work. The DPR approach was similar to the traditional panel-based approach. Additional studies with larger numbers of participants are planned.

## References

- Lewis J, Schneegans S, Straza T. *UNESCO Science Report: The Race Against Time for Smarter Development*. Vol 2021. UNESCO Publishing; 2021.
- Merrifield MR, Saari DG. Telescope time without tears: a distributed approach to peer review. *Astronomy and Geophysics*. 2009;50(4):4.16-4.20. doi:10.1111/j.1468-4004.2009.50416.x
- Andersen M, Chiboucas K, Geball T, et al. The Gemini Fast Turnaround program. In: *Abstracts of the 233rd AAS Meeting; January 6-10, 2019; Seattle, Washington*. Vol 233. American Astronomical Society; 2019:761

<sup>1</sup>Michigan State University, East Lansing, MI, USA, wkerzend@msu.edu; <sup>2</sup>European Southern Observatory, Garching bei München, Germany; <sup>3</sup>University of Pittsburgh, Pittsburgh, PA, USA; <sup>4</sup>University of Vienna, Vienna, Austria; <sup>5</sup>New York University, New York, NY, USA

**Conflict of Interest Disclosures** Wolfgang Kerzendorf is part of New York University, and the SNYU group is supported by US National Science Foundation CAREER awards AST-1352405 and AST-1413260. He was also supported by a European Southern Observatory (ESO) Fellowship and the Excellence Cluster Universe, Technische Universität München, for part of this work. Glenn van de Ven acknowledges funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program with grant agreement 724857 (consolidator grant ArcheoDyn).

**Funding/Support** Wolfgang Kerzendorf was funded by the ESO, New York University, and Michigan State University. The other authors were funded by their respective institutions.

**Acknowledgments** We thank the 167 volunteers who participated in the distributed peer review (DPR) experiment for their work and enthusiasm. We also thank M. Kissler-Patig for promoting the DPR experiment following his experience at Gemini; ESO’s director general, X. Barçons, and director for science, R. Ivison, for their support; and H. Schütze for several suggestions on the natural language processing. We thank J. Linnemann for help with some of the statistics tests. We acknowledge the help of Michael Berkwits and Leah Dickstein for the specific questions of norms in the medical community. Wolfgang Kerzendorf thanks the Flatiron Institute.

**Disclaimer** This abstract is the result of independent research and is not to be considered as expressing the position of the ESO on proposal review and telescope time allocation procedures and policies.

## The Gap Between Reviewers’ Recommendations and Editorial Decisions in a Medical Education Journal

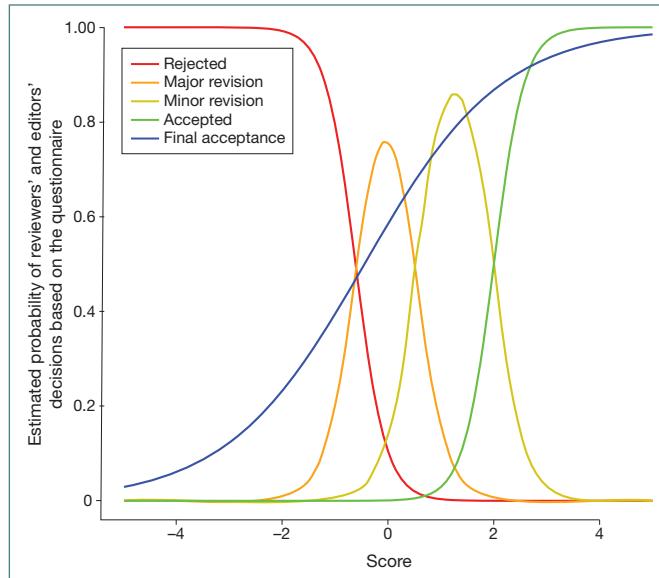
José J. Naveja,<sup>1,2</sup> Daniel Morales-Castillo,<sup>1</sup> Teresa Fortoul,<sup>1</sup> Melchor Sánchez-Mendiola,<sup>1</sup> Carlos Gutiérrez-Cirlos<sup>1</sup>

**Objective** To assess the items in a questionnaire and identify critical points that might be associated with manuscript final decisions. Usually, peer reviewers’ input is in the form of a series of free-format suggestions for the manuscript under consideration.<sup>1</sup> Typically, journals ask reviewers a few questions about the overall quality of the work.<sup>2</sup> The validity evidence and psychometric properties of this type of instrument in the editorial process are not commonly evaluated, and the goal of the study was to assess them in the Mexican journal *Investigación en Educación Médica*.

**Design** In this cohort and instrument analysis, during the peer review process of the journal, a 15-item questionnaire was used, with dichotomous response referring to the manuscript subsections: abstract (1 item), introduction (3 items), methods (2 items), results (1 item), discussion and/or conclusions (5 items), and general evaluation (3 items). The questionnaires for manuscripts that finished peer review from January 2020 to December 2021 were analyzed. Cronbach  $\alpha$  was used to measure reliability. An item-response-theory model fit was used to identify the best items for discriminating publishable manuscripts.

**Results** A total of 169 reviewer reports were collected from 85 manuscripts that underwent peer review with a final editorial decision (mean, 1.99 reviews per manuscript). Missing data were found in 2.7% of the responses. The Cronbach  $\alpha$  score for reliability was 0.86, when considering the questionnaire and including the reviewer’s final recommendation, and the Cronbach  $\alpha$  score for the editor’s decision was 0.88. A 2-parameter item-response-theory model fitted the data well (root mean square error of approximation, 0.05; Tucker-Lewis index, 0.97). This means that a latent variable (ie, the overall reviewers’ impression of the manuscript) seemed to be associated with the response to the items. The model allowed the integration of all responses into a combined score, and some items were more likely to be marked false than others when the overall score was low (**Figure 22**). The variability identified in the estimated item

**Figure 22. Comparison of Response Functions for Reviewer Recommendations and Editor's Final Decision**



Comparison of response functions for reviewer recommendations (rejected, major and minor revision, accepted) and editor's final decision (final acceptance) in the item response theory model. The score (x-axis) depends on the response pattern in the survey. Note that the editor's decisions catch up slowly with reviewers' opinions. Numbers on the y-axis represent the probability that any given recommendation category is met by the reviewer (rejected, major and minor revisions, accepted) or the editor (final acceptance), assuming that the other answers in the questionnaire are known.

discrimination parameters showed that some were more informative than others. The highest discrimination was observed in items related to the validity of inferences and conclusions. The correlation between instrument scores and reviewers' recommendations was high ( $r = 0.88$ ) but translated poorly into editors' final decisions. For instance, the discrimination parameter for the editors' decisions was relatively small (Cronbach  $\alpha$  score, 0.77).

**Conclusions** The study adds internal validity evidence about an instrument that provides editors with an overview of the manuscript. The model transformed the questionnaire responses into a score that better captured the reviewers' impressions. Discrepancies between reviewers and editors cannot be avoided.<sup>3</sup> The reviewers' questionnaire is a useful starting point when reaching an editorial decision after peer review. Journals that include a brief questionnaire for assessing the manuscripts can benefit from analyzing these responses. A general questionnaire cannot replace the specific comments written by reviewers for the final editorial decision.

## References

1. Sun M. Peer review comes under peer review. *Science*. 1989;244(4907):910-912. doi:10.1126/science.2727683
2. Parrish E. Peer review process. *Perspect Psychiatr Care*. 2021;57(1):7-8. doi:10.1111/ppc.12724
3. Snell L, Spencer J. Reviewers' perceptions of the peer review process for a medical education journal. *Med Educ*. 2005;39(1):90-97. doi:10.1111/j.1365-2929.2004.02026.x

<sup>1</sup>Faculty of Medicine, National Autonomous University of México, Mexico City, Mexico, cirlas@hotmail.com; <sup>2</sup>Johannes Gutenberg University, Mainz, Germany

**Conflict of Interest Disclosures** None reported.

**Additional Information** Melchor Sánchez-Mendiola is a co-corresponding author.

## Preprints

### Downstream Retraction of Preprinted Research in the Life and Medical Sciences

Michele Avissar-Whiting<sup>1</sup>

**Objective** Retractions have been on the rise in the life and clinical sciences in the last decade.<sup>1</sup> In this same period, there has been a greater than 10-fold increase in the posting of preprints by researchers in these fields.<sup>2</sup> These developments have introduced challenges with respect to the back propagation of events such as retractions that occur on the journal-published version. The aim of this study was to understand the extent of this problem among servers that routinely link their preprints to their corollary versions published in journals.

**Design** To present a snapshot of the current state of downstream retractions of articles preprinted in 3 large preprint servers (Research Square, bioRxiv, and medRxiv), the DOIs of the journal-published versions linked to preprints were matched to entries in the Retraction Watch database. The analysis covered November 2013 to November 2021. Preprints with downstream retractions were checked for (1) a notice of the retraction and (2) a notice of withdrawal of the preprint. The times from preprinting to publication and publication to retraction were calculated, and the types of misconduct (mistreatment of research subjects, falsification and fabrication of data, and piracy and plagiarism) were categorized for each retraction.

**Results** A total of 30 retractions were identified, representing only 0.03% of all content with journal links posted on these servers. Of these, 11 (36.7%) were clearly noted by the preprint servers; however, the existence of a preprint was only acknowledged by the retracting journal in 1 instance (3.3%). The mean time from publication to retraction was 9.2 months, notably lower than the mean of 22.3 months for articles overall. In 20 of 30 cases (66.7%), retractions downstream of preprints were due, at least in part, to some form of research misconduct. In 18 of 30 cases (60%), the nature of the retraction suggested that the conclusions were no longer reliable.

**Conclusions** As adoption of preprints continues to grow, serious consideration should be given to ensuring that preprints are digitally connected with associated publications and building reliable mechanisms for propagating critical updates. It is incumbent on preprint servers, journals, and the systems that connect them to address these issues before their scale becomes untenable.

## References

1. Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased? *PLoS ONE*. 2013;8:e68397. doi:10.1371/journal.pone.0068397
2. Vasconcelos GL, Cordeiro LP, Duarte-Filho GC, Brum AA. Modeling the epidemic growth of preprints on COVID-19 and SARS-CoV-2. *Front Phys*. Published online March 23, 2021. doi:10.3389/fphy.2021.603502

<sup>1</sup>Research Square Company, Durham, NC, USA, michavissar@gmail.com

**Conflict of Interest Disclosures:** At the time of this study, Michele Avissar-Whiting was the editor in chief of the Research Square preprint platform.

## Quality of Trials

### A Screening Checklist to Assess Data Integrity and Fabrication in Randomized Clinical Trials

Ben W. Mol,<sup>1,2</sup> Shimonla Lai,<sup>1</sup> Ayesha Rahim,<sup>1</sup> Wentao Li<sup>1</sup>

**Objective** Because randomized clinical trials (RCTs) inform clinical guidelines and can directly influence patient care, it is important to ensure that the data behind their conclusions are trustworthy. We aimed to develop a checklist to screen RCTs for possible data fabrication at submission or prior to inclusion in meta-analyses.

**Design** The development of this screening tool was adapted from the 4-stage approach proposed by Moher et al<sup>1</sup> for reporting guidelines, including defining the scope, reviewing the evidence base, suggesting a list of items from piloting, and holding a consensus meeting as part of a Delphi method. The

initial checklist was set up by a smaller core group based on the authors' experience assessing problematic RCTs for several years. The checklist was then piloted in a Delphi panel of 20 stakeholders, including clinicians, reviewers, journal editors, and evidence synthesis specialists. Using a set of 15 articles, 8 of which were known to have fabricated data, each member was asked to score 3 articles with the checklist. Results were then discussed in 2 Delphi sessions.

**Results** The screening checklist had 7 domains and is detailed in **Table 48**. The group proposed that a positive screen be defined by 2 or more items present.

**Conclusions** This is the first checklist developed in a formal process to detect possible data fabrication in RCTs. If a study is assessed and found to be suspicious, reviewers can consider a more thorough investigation into the data integrity issues identified, including assessment of original data. This checklist may help editors, publishers, and researchers to screen for data fabrication in RCTs in an objective manner.

## Reference

1. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLOS Medicine*. 2010;7(2):e1000217. doi.org/10.1371/journal.pmed.1000217

<sup>1</sup>Department of Obstetrics and Gynaecology, Monash University, Clayton, Australia, ben.mol@monash.edu; <sup>2</sup>Aberdeen Centre for Women's Health Research, Institute of Applied Health Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK

**Conflict of Interest Disclosures** Ben W. Mol is supported by the National Health and Medical Research Council (grant GNT1176437), reports consultancy for ObsEva and Merck, and receives travel support from Merck. No other disclosures were reported.

**Table 48. Screening Checklist for Data Fabrication in Randomized Clinical Trials**

Domain	Description
Governance	Absent/retrospective registration
	Discrepancy trial registration/publication
	Absent/vague description of ethics
Author group	≤3 Authors; low author to study size ratio
	Authors with retracted articles; large number of randomized clinical trials by 1 author in short time
Plausibility of intervention usage	Sealed envelopes in placebo-controlled trial
	Implausible use of placebo (eg, 2 interventions but only 1 placebo in 3-arm trial)
Time frame	Implausibly fast recruitment
	Fast submission
	No time to allow follow-up
Dropout rates	Zero follow-up
	Rounded number in each group
Baseline characteristics	Few baseline characteristics
	Implausible patient characteristics judging from common sense, literature, or local data
	Perfect balance or significant/large differences

## Registries and Repositories

### A Comprehensive Assessment of Changes to Prespecified Trial Outcomes, Including Historical Registry Records

Martin R. Holst,<sup>1</sup> Martin Haslberger,<sup>1</sup> Daniel Strech,<sup>1</sup> Lars G. Hemkens,<sup>2,3,4</sup> Benjamin G. Carlisle<sup>1</sup>

**Objective** Preregistration plays a key role in reducing certain biases in clinical trial conduct and reporting. Most studies on selective reporting have assessed discrepancies between the latest registry entry and the corresponding publication. However, ClinicalTrials.gov and the German Clinical Trials Register (DRKS) allow entries to be updated after initial registration, which allows researchers to add or change crucial details. Because studies look at only the latest registry entry, these later changes are not directly visible, which constitutes another source of possible reporting bias if larger changes go unreported. By assessing the entire audit trail, which was not easily accessible for mass download and analysis until recently, this study was a comprehensive analysis of outcome changes made within registry entries between key study time points and in corresponding results

publications and analyzed factors associated with these changes.

**Design** Based on an existing data set of all trials conducted at German university medical centers between 2009 and 2017 and registered in DRKS or ClinicalTrials.gov that had published results,<sup>1,2</sup> all historical registry entries were obtained using the R package cthist.<sup>3</sup> Historical trial registration records from the 2 databases were extracted and semiautomatically evaluated. For each trial, changes to primary outcomes between key study time points in the registry (first patient inclusion, completion, publication, and latest available version) and publication were extracted. These changes were then classified according to their severity, and associations with a number of candidate predictors of outcome changes at different trial stages were analyzed.

**Results** Of 1747 included trials, 592 trials (33.9%) had an outcome change of some kind within the registration after study start. Analyses of outcomes in publications and the nature and severity of outcome changes are ongoing, and results will be presented at the conference. The study also examined whether and how outcome changes were reported in results publications and presented factors associated with outcome changes at different trial stages.

**Conclusions** A large proportion of clinical trials exhibited changes to prespecified outcomes within the registry after study start, and the nature of these changes is important to know for the integrity of the scientific process. This analysis provided further insight into outcome registration and reporting practices. Using methods reported in this study, peer reviewers, editors, readers, and metaresearchers may be able to assess the registration quality of a clinical trial.

## References

1. Riedel N, Wieschowski S, Bruckner T, et al. Results dissemination from completed clinical trials conducted at German university medical centers remained delayed and incomplete: the 2014–2017 cohort. *J Clin Epidemiol*. 2022;144:1-7. doi:10.1016/j.jclinepi.2021.12.012
2. Wieschowski S, Riedel N, Wollmann K, et al. Result dissemination from clinical trials conducted at German university medical centers was delayed and incomplete. *J Clin Epidemiol*. 2019;115:37-45. doi:10.1016/j.jclinepi.2019.06.002
3. Carlisle BG. Analysis of clinical trial registry entry histories using the novel R package cthist. *medRxiv*. Preprint posted online January 21, 2022. doi:10.1101/2022.01.20.22269538

<sup>1</sup>Berlin Institute of Health at Charité—Universitätsmedizin Berlin, QUEST Center for Responsible Research, Berlin, Germany; martin.haslberger@bih-charite.de; <sup>2</sup>Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland;

<sup>3</sup>Berlin Institute of Health at Charité—Universitätsmedizin Berlin, QUEST Center for Responsible Research, Meta-Research Innovation Center Berlin, Berlin, Germany; <sup>4</sup>Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, USA

**Conflict of Interest Disclosures** The authors declare no direct conflicts of interest. Daniel Streich is a member of the Sanofi

Advisory Bioethics Committee and receives an honorarium for his contribution to meetings.

**Funding/Support** The project is funded from QUEST departmental resources.

**Additional Information** This study protocol was preregistered (<https://osf.io/t3qva>). Martin R. Holst and Benjamin G. Carlisle are co-corresponding authors.

## Reporting Guidelines

### A Mapping Review of Comments on SPIRIT 2013 and CONSORT 2010 Reporting Guidelines for Reporting Randomized Trials

Camilla Hansen Nejstgaard,<sup>1,2</sup> Isabelle Boutron,<sup>3,4</sup> An-Wen Chan,<sup>5</sup> Ryan Chow,<sup>6</sup> Sally Hopewell,<sup>7</sup> Mouayad Masalkhi,<sup>8</sup> David Moher,<sup>9,10</sup> Kenneth F. Schulz,<sup>11</sup> Nathan A. Shlobin,<sup>12</sup> Lasse Østengaard,<sup>1,2,13</sup> Asbjørn Hróbjartsson<sup>1,2</sup>

**Objective** When appropriately designed, conducted, and reported, randomized trials provide trustworthy assessments of the effects of health care interventions. The SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) guideline was developed in 2013 to facilitate complete and transparent reporting in trial protocols. Similarly, the CONSORT (Consolidated Standards of Reporting Trials) guideline was developed in 1996, with subsequent updates in 2001 and 2010, to facilitate complete and transparent reporting of trial methods and results. Periodic guideline updates are essential to reflect evolving trial methodologies and experiences with implementation. The objective of this study was to identify, summarize, and analyze comments on both guidelines, with special emphasis on suggestions for guideline modifications.

**Design** This mapping review (reported in accordance with PRISMA-ScR [Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews]), included documents (eg, empirical studies and letters) written in English and published after 2010 with explicit comments on SPIRIT 2013 or CONSORT 2010. Three bibliographic databases (Embase, Medline to June 2020, and Web of Science to April 2022) and other sources (eg, Google Scholar, BMC Blog Network, *The BMJ* rapid responses, and proceedings from Cochrane Colloquia) were searched. Two authors independently assessed documents for eligibility and extracted data on basic characteristics and exact wording of the main comments. Comments were categorized as suggestions for modification to the wording of existing guideline item, suggestions for new item, or reflections on challenges or strengths. The SPIRIT or CONSORT topic addressed (eg, methods or results) and the item number were noted. Suggestions were summarized and categorized into those that were directly linked to empirical investigations, were continuations of previous methodological discussions, or reflected new methodological developments.

**Results** Searches identified 7324 records, of which 82 documents with 99 comments were included. In total, 36

comments suggested modifying existing guideline items. The section most commented on was the participant flow section of CONSORT (eg, add numbers on nonrandomized screened participants, n = 8). There were 37 suggestions for new items in both SPIRIT and CONSORT. Multiple comments addressed intervention (eg, add content on cointerventions, n = 7), blinding (eg, add content on risk of unblinding, n = 7), participant flow (eg, add content on missing data, n = 5), and statistical methods (eg, add content on blinding of statisticians, n = 4). Thirty-six of the suggestions (49%) were directly linked to empirical investigations. Six of the suggestions (8%) were continuations of previous methodological discussions, and 4 suggestions (5%) reflected new methodological developments.

**Conclusions** Ninety-nine comments on SPIRIT 2013 and CONSORT 2010 were identified. The suggestions for modifying the wording or adding new items were often related to participant flow, intervention, blinding, and statistical methods. Approximately half of the suggestions were directly linked to empirical investigations. The issues raised may provide helpful context to authors, peer reviewers, editors, and readers of trials using SPIRIT 2013 and CONSORT 2010 and inform future updates to these guidelines.

<sup>1</sup>Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark, cahansen@health.sdu.dk; <sup>2</sup>Open Patient Data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark; <sup>3</sup>Université de Paris, Centre of Research in Epidemiology and Statistics (CRESS), Inserm, France; <sup>4</sup>Cochrane France, Paris, France; <sup>5</sup>Department of Medicine, Women's College Research Institute, University of Toronto, Canada; <sup>6</sup>Faculty of Medicine, University of Ottawa, Canada; <sup>7</sup>Oxford Clinical Trials Research Unit/Centre for Statistics in Medicine, University of Oxford, UK; <sup>8</sup>School of Medicine, University College Dublin, Dublin, Ireland; <sup>9</sup>Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Canada; <sup>10</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Canada; <sup>11</sup>Department of Obstetrics and Gynecology, School of Medicine, The University of North Carolina at Chapel Hill; <sup>12</sup>Department of Neurological Surgery, Northwestern University Feinberg School of Medicine, Chicago, Illinois; <sup>13</sup>University Library of Southern Denmark, Odense, Denmark

**Conflict of Interest Disclosures:** David Moher is an associate director and Isabelle Boutron and An-Wen Chan are advisory board members of the International Congress on Peer Review and Scientific Publication but were not involved in the review or decision of this abstract. No other disclosures were reported.

### Development of the Accurate Consensus Reporting Document (ACCORD) Reporting Guideline

Patricia Logullo,<sup>1</sup> Esther J. van Zuuren,<sup>2</sup> A. Pali S. Hungin,<sup>3</sup> Christopher C. Winchester,<sup>4</sup> David Tovey,<sup>5</sup> Ellen L. Hughes,<sup>6</sup> Keith H. Goldman,<sup>7</sup> Niall Harrison,<sup>8</sup> William Gattrell<sup>9</sup>

**Objective** Consensus methodologies are widely used to harness expert knowledge for decision-making in areas of uncertainty. While specific guidance is available on conducting and reporting Delphi initiatives in palliative care,<sup>1</sup>

there remains a need for broader reporting guidelines. The ACCORD (Accurate Consensus Reporting Document<sup>2</sup>) initiative will develop guidance for transparent and complete reporting of consensus-building methodologies in biomedical research and clinical practice. This abstract reports findings from a systematic literature review that will inform consensus on checklist items for ACCORD.

**Design** Studies, reviews, and guidance documents addressing the quality of reporting of consensus methods in biomedicine or clinical practice were eligible for inclusion. Reports of consensus methods that did not comment on reporting quality were excluded. Searches were conducted with no limits by year or language. Identified articles were retrieved and assessed for eligibility using Rayyan by 4 evaluators working independently; discrepancies were reconciled by discussion. The search process started on January 7, 2022; the assessment period was completed on February 7, 2022. Data extraction was done using Covidence, and potential checklist items were generated.

**Results** Overall, 2736 references were identified: 2599 articles and documents (Web of Science, 1775; MEDLINE [Web of Science], 1501 [202 unique]; PubMed, 375 [219 unique]; MEDLINE [OVID], 641 [174 unique]; Embase, 331 [66 unique]; Cochrane, 131 [77 unique]; Emcare, 179 [29 unique]; Academic Search Premier, 280 [23 unique]; and PsycINFO, 173 [34 unique]) and 137 meeting abstracts (Web of Science, 14; Embase, 99 [90 unique]; Cochrane, 36 [33 unique]). In all, 54 publications were selected for full-text review; 18 met the eligibility criteria. Most studies acknowledged that reporting quality of consensus initiatives could be improved. The most discussed items included panel composition and consensus definition and thresholds (**Table 49**). Public and patient involvement and roles of the steering committee and chair(s) were among the least addressed.

**Conclusions** Most identified studies acknowledged the need to improve reporting quality of consensus methodologies. ACCORD aims to set a standard for reporting consensus initiatives, improving their transparency and making it easier to critically appraise the methods used to develop consensus recommendations.

### References

- Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and reporting delphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. *Palliat Med* 2017;31:684-706. doi:10.1177/0269216317690685
- Gattrell WT, Hungin AP, Winchester CC, et al. ACCORD guideline for reporting consensus-based methods in biomedical research and clinical practice: a study protocol. *Res Integr Peer Rev*. 2022;7(3): doi:10.1186/s41073-022-00122-0

<sup>1</sup>Centre for Statistics in Medicine (CSM), University of Oxford, and EQUATOR Network UK Centre, Oxford, UK, patricia.logullo@ndorms.ox.ac.uk; <sup>2</sup>Leiden University Medical Centre, Leiden,

**Table 49. Reporting Items Discussed**

Reporting item	No. (%) of studies including guidance
<b>Background</b>	
1.1 Rationale for choosing a consensus method over other methods	4 (22.2)
1.2 Clearly defined objective	6 (33.3)
<b>Methods</b>	
2.1 Review of existing evidence informing consensus study	5 (22.8)
2.2 Inclusion and exclusion criteria of the literature search	3 (16.7)
2.3 Composition of the panel	16 (88.9)
2.4 Public patient involvement	0
2.5 Panel recruitment	4 (22.2)
2.6 Defining consensus and the threshold for achieving consensus	13 (72.2)
2.7 Decision of item approval	3 (16.7)
2.8 Number of voting rounds	10 (55.6)
2.9 Rationale for number of voting rounds	8 (44.4)
2.10 Time between voting rounds	1 (5.6)
2.11 Additional methods used alongside consensus	2 (11.1)
2.12 Software or tools used for voting	1 (5.6)
2.13 Anonymity of panelists and how this was maintained	7 (38.9)
2.14 Feedback to panelists at the end of each round	11 (6.1)
2.15 Synthesis/analysis of responses after voting rounds	5 (27.8)
2.16 Pilot testing of study material/instruments	3 (16.7)
2.17 Role of the steering committee/chair/co-chair/facilitator	0
2.18 Conflict of interest or funding received	4 (22.2)
2.19 Measures to avoid influence by conflict of interest	1 (5.6)
<b>Results</b>	
3.1 Results of the literature search	1 (5.6)
3.2 Number of studies found as supporting evidence	0
3.3 Response rates per voting round	5 (27.8)
3.4 Results shared with respondents	9 (50.0)
3.5 Dropped items	5 (27.8)
3.6 Collection, synthesis, and comments from panelists	5 (27.8)
3.7 Final list of items (eg, for guideline or reporting guideline)	4 (22.2)
<b>Discussion</b>	
4.1 Limitations and strengths of the study	5 (27.8)
4.2 Applicability, generalizability, reproducibility	3 (16.7)

Netherlands; <sup>3</sup>University of Newcastle, Newcastle upon Tyne, UK; <sup>4</sup>Oxford PharmaGenesis, Oxford, UK; <sup>5</sup>*Journal of Clinical Epidemiology*, London, UK; <sup>6</sup>Sciwright Limited, Somerset, UK; <sup>7</sup>AbbVie, North Chicago, IL, USA; <sup>8</sup>Ogilvy Health, London, UK; <sup>9</sup>Ipsen, Milton Park, UK

**Conflict of Interest Disclosures** Patricia Logullo is a member of the UK EQUATOR Centre, Oxford, UK, an organization that promotes the use of reporting guidelines, many of which are developed using consensus methods, and she is personally involved in the development of other reporting guidelines.

A. Pali S. Hungin worked with Reckitt Benckiser in the last 5 years for the development of the definitions and management of gastroesophageal reflux disease. Christophe C. Winchester is an employee, director, and shareholder of Oxford PharmaGenesis, Oxford, UK; a director of Oxford Health Policy Forum CIC; a trustee of the Friends of the National Library of Medicine; and an associate fellow of Green Templeton College. Ellen L. Hughes has worked with Ogilvy Health UK on consensus projects. William Gattrell was an employee of Ipsen, Oxford, UK, at the time of the analysis. He is now an employee of Bristol Myers Squibb. No other disclosures were reported. No authors were reimbursed for participating in the initiative.

## Guiding Principles for Updating Reporting Guidelines: A Qualitative Analysis

Patrick M. Bossuyt,<sup>1</sup> Constantine A. Gatsonis,<sup>2</sup> Jérémie F. Cohen<sup>3</sup>

**Objective** The Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline for studies to evaluate the diagnostic accuracy studies of medical tests was published in 2003. The STARD reporting guideline was developed to “improve the accuracy and completeness of reporting ... to allow readers to assess the potential for bias in the study and to evaluate its generalisability.” An update of the reporting guideline appeared in 2015. The STARD executive team has started making preparations for another update. In doing so, the aim was to explore whether other recent reporting guidelines emphasized other guiding principles, beyond internal validity (risk of bias) and external validity (applicability). Such principles could be concerns about conflicts of interest, or equity, or diversity.

**Design** Based on a search of the published literature and the EQUATOR network, a convenience sample of 24 reporting guidelines was assembled. Explicitly mentioned guiding principles were extracted and scrutinized in these reporting guidelines to identify other reasons for including or emphasizing specific elements in study reports. Using an inductive approach, these principles and arguments were organized into a smaller number of categories.

**Results** Although the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline and other early reporting guidelines have always emphasized the need for methodologic rigor, reporting guidelines are, and have always been, based on a range of principles, including more than internal validity and risk of bias. These principles include, but are not limited to, the following factors: identifiability in literature searches, clarity, integrity (competing interests and sources of funding), open science (accessibility), and general scientific principles (prior literature, hypotheses, limitations). More recent guidelines include additional considerations, such as ethics, equity and fairness, sex and gender considerations, reducing research waste, and curbing selective reporting and spin in reporting and interpretation.

**Conclusions** Reporting guidelines are based on a growing range of guiding principles. This situation poses challenges to groups interested in developing or updating a reporting guideline, who will have to balance completeness and

complexity, presenting specific items or including generic ones, and, sometimes, deciding on being complete versus being helpful, for authors, reviewers, and editors.

<sup>1</sup>Department of Epidemiology & Data Science, Amsterdam University Medical Centers, University of Amsterdam, the Netherlands, p.m.bossuyt@amsterdamumc.nl; <sup>2</sup>Center for Statistical Sciences, Department of Biostatistics, Brown University, Providence, RI, USA; <sup>3</sup>Centre of Research in Epidemiology and Statistics (CRESS), Obstetrical, Perinatal and Pediatric Epidemiology Research Team, INSERM, Université de Paris, France

**Conflict of Interest Disclosures** All authors contributed to the development of STARD 2015. Patrick M. Bossuyt is a member of the Peer Review Congress advisory board but was not involved in the review or decision of this abstract.

## Retractions

### A Survey of Approaches Taken by Medical Libraries to Educate Users About Retracted Biomedical Publications

Peiling Wang,<sup>1</sup> Lisa Ennis<sup>2</sup>

**Objective** A growing rate of retractions in biomedical literature is a threat to the scientific publishing ecosystem. On January 7, 2022, the Web of Science had 7776 retracted biomedical articles published in 2000–2022, and 113,931 articles cited these retracted articles. The continued use of these retracted publications as valid scientific findings is a serious problem.<sup>1–3</sup> Medical libraries can play a significant role in their user education and support programs to increase awareness and provide tools. This study aims to investigate

how medical schools' libraries teach users about retracted literature.

**Design** Currently, there are 154 MD and 38 DO programs in the USA. An online survey was distributed to the Medical Library Association's lists of 12 Chapters and the Osteopathic Libraries Caucus (88 members). The 9-item questionnaire collected data on the offered degrees, the library programs' topics, instructional modes, and free-text comments as well as the importance of covering retracted literature.

**Results** A total of 85 questionnaires were received, of which 38 responses were completed. Some of the incomplete questionnaires were from librarians in university or hospital-affiliated medical libraries. The 38 completed responses were from 22 medical schools within public universities or standalone (58%) and 16 medical schools within private universities or standalone private (42%). These programs offer 28 MD and 13 DO degrees. The topics covered how to cite (74%), copyright (71%), plagiarism (43%), retractions (37%), and 6 other topics filled by 6 respondents (funding supports, use of image, predatory journals, predatory publications, ghost authorship, and Salami factor). The delivery modes included how-to guides (82%), sessions in courses (68%), faculty workshops (47%), video tutorials (37%), and lunch-and-learn sessions for students or faculty (29%). The mean (SD) importance of covering retracted literature was 3.74 (1.00) on a 5-point Likert scale. For the libraries having taught or planning to cover retracted literature, the scope and levels varied (**Table 50**). For the 63% of libraries not yet including retractions in instructions, reasons included “would love to but it's hard enough to get

**Table 50. Useful Free-Text Comments on Covering Retracted Literature in Educational Programs**

Representative comments	Instruction modes	How retractions were covered
We highlight it with databases like Retraction Watch, and integrations like LibKey. It's really just to let them know that it is a possibility and to be on the lookout. You should always be a little skeptical and avoid relying just on 1 source.	<ul style="list-style-type: none"> <li>• Faculty workshops</li> <li>• Sessions in courses</li> <li>• Lunch-and-learn sessions</li> <li>• How-to guides</li> </ul>	<ul style="list-style-type: none"> <li>• Awareness of retractions</li> <li>• Need to check multiple sources</li> <li>• Database: Retraction Watch</li> <li>• Tools (LibKey)</li> </ul>
Try to introduce people to the complexities of this topic and offer some strategies to identify retractions.	<ul style="list-style-type: none"> <li>• Faculty workshops</li> <li>• Sessions in courses</li> <li>• How-to guides</li> </ul>	<ul style="list-style-type: none"> <li>• Awareness of retractions (complexity)</li> <li>• How to identify retracted articles</li> </ul>
Students are introduced to the idea of how much literature is out there and how much has been retracted and why. We also discuss how to identify and select reliable journals.	<ul style="list-style-type: none"> <li>• Sessions in courses</li> <li>• How-to guides</li> </ul>	<ul style="list-style-type: none"> <li>• Publication growth</li> <li>• Awareness of retractions (reasons)</li> <li>• How to select journals</li> </ul>
Checking for article retractions would be a step. For example, EndNote 20 has a retraction alert feature that might prove helpful.	<ul style="list-style-type: none"> <li>• Evidence synthesis project</li> <li>• How-to guides</li> </ul>	<ul style="list-style-type: none"> <li>• Awareness of retractions (checking)</li> <li>• Tools (EndNote 20)</li> </ul>
Sometimes, we mention it when we talk about systematic review projects: check that included studies haven't been retracted.	<ul style="list-style-type: none"> <li>• Faculty workshops</li> <li>• Sessions in courses</li> <li>• How-to guides</li> </ul>	<ul style="list-style-type: none"> <li>• Awareness of retractions (systematic review)</li> </ul>
We do use examples, but they don't go in depth. It's probably increased in importance because of COVID-19.	<ul style="list-style-type: none"> <li>• Sessions in courses</li> <li>• Video tutorials</li> <li>• How-to guides</li> </ul>	<ul style="list-style-type: none"> <li>• Awareness of retractions (not in depth)</li> </ul>
Thanks for raising this important topic.	<ul style="list-style-type: none"> <li>• Sessions in courses</li> </ul>	<ul style="list-style-type: none"> <li>• Needs for covering</li> </ul>
This has not really been addressed yet, but we believe it is beginning to be an issue that needs to be addressed.	<ul style="list-style-type: none"> <li>• Sessions in courses</li> <li>• How-to guides</li> </ul>	<ul style="list-style-type: none"> <li>• Needs for covering</li> </ul>
This is a very important topic. I plan on discussing it after this survey.	<ul style="list-style-type: none"> <li>• Faculty workshops</li> <li>• Video tutorials</li> </ul>	<ul style="list-style-type: none"> <li>• Plans for covering</li> </ul>
I haven't considered it, but it is important. It could be easier to include it in the faculty research workshop series.	<ul style="list-style-type: none"> <li>• Faculty workshops</li> </ul>	<ul style="list-style-type: none"> <li>• Plans for covering</li> </ul>
I'm now considering adding retracted biomedical literature to sessions.	<ul style="list-style-type: none"> <li>• Sessions in courses</li> </ul>	<ul style="list-style-type: none"> <li>• Plans for covering</li> </ul>

into how to use the library resources" (perceived time limit); "important to include, but I need to learn more before I can develop instruction beyond 'here's how to check"'; and "what would we teach?" (knowledge gap).

**Conclusions** The preliminary results suggest that the coverage of retracted literature by library education ranged from ad hoc to substantial. Librarians are aware of the needs to discuss the retracted literature in-depth but are limited by the allotted time for library instruction or knowledge gaps. To close the knowledge gap on retracted literature in the biomedical ecosystem, Medical Library Association competencies should address this competence. Effective tools for tracking and alerting the retracted literature can help faculty and students. Given the low response rate, a revised survey should include all medical libraries beyond medical schools' libraries.

## References

1. Pfeifer MP, Snodgrass GL. The continued use of retracted, invalid scientific literature. *JAMA*. 1990;263(10):1420-1423.
2. Wang P, Su J. Expert-recommended biomedical journal articles: their retractions or corrections, and post-retraction citing. *J Inf Sci*. Published online February 2, 2022. doi:10.1177/01655515221074329
3. Wang P, McCullough L, Su J. Continued use of retracted publications: implications for scientific publishing and information systems. iConference. Published online February 2, 2022. [https://trace.tennessee.edu/utk\\_infosciepubs/462/](https://trace.tennessee.edu/utk_infosciepubs/462/)

<sup>1</sup>School of Information Sciences, University of Tennessee-Knoxville, Knoxville, TN, USA, peilingw@utk.edu; <sup>2</sup>Library & Learning Resources, Alabama College of Osteopathic Medicine, AL, USA

**Conflict of Interest Disclosures** None reported.

**Additional Information** Peiling Wang and Lisa Ennis are co-first authors.

## SATURDAY, SEPTEMBER 10

### Artificial Intelligence

#### Rejection Rates for Manuscripts Uploaded to an Artificial Intelligence–Driven Precheck Tool Compared With Manuscripts That Did Not Undergo a Precheck at a Multidisciplinary Medical Journal

Duncan A. MacRae,<sup>1</sup> Abhishek Sudra,<sup>2</sup> Kara Hamilton<sup>1</sup>

**Objective** Online precheck tools identify common errors in grammar and formatting and are intended to help authors identify missing declarations and common language issues prior to first submission. The purpose of this study was to evaluate the use of an artificial intelligence–driven precheck tool and to examine the resulting association with initial rejection rates.

**Design** This cohort study involved original research manuscripts submitted to *Medicine*, an open access

multidisciplinary medical journal, during a 7-month period from June 2021 to January 2022. Prior to submission, authors were encouraged to upload their manuscript to an online artificial intelligence–driven precheck tool, which understands the precise meaning of phrases within a document and automatically captures both semantic and syntactic variations. The tool is configured to check for language and grammar quality as well as the presence of ethics statements, conflicts of interest declarations, and adherence to word count limits. The precheck tool offers 2 levels of feedback: a free basic report, which summarizes issues that the system suggests should be addressed prior to submission, and a premium check (costing US \$29), which provides the author with a downloadable Word document containing all suggested changes in detail. Authors were not mandated to use the precheck tool, and the choice to purchase the premium report was entirely at the author's discretion. The resulting report was provided to the authors so that changes could be made prior to submission. The journal editors did not receive a copy of the report. All manuscripts were also subjected to a technical check carried out by the editorial office prior to the assignment of editors or reviewers. Articles uploaded to the precheck tool platform were then crosschecked against all articles submitted to the journal's submission platform, allowing the journal to compare the proportions initially rejected (ie, decisions made prior to undergoing peer review) amongst the 3 distinct groups.

**Results** Among 7904 submitted manuscripts, author selections for the 3 groups of manuscripts (no precheck, basic precheck, and premium precheck) and numbers initially rejected are detailed in **Table 51**. Among manuscripts in the no precheck group, 2073 of 6062 (34.2%) were rejected following technical check compared with 333 of 1661 (20.1%) in the basic precheck group and 13 of 181 (7.3%) in the premium precheck group. Overall, 15.4% fewer manuscripts that underwent prechecking were rejected compared with those that underwent no prechecking (346 of 1842 [18.8%] vs 2073 of 6062 [34.2%]).

**Conclusions** The use of a precheck tool to assist authors in identifying language errors and missing manuscript elements prior to submission was associated with a decrease in initial manuscript rejections (**Table 51**).

<sup>1</sup>Wolters Kluwer Health, Philadelphia, PA, USA, duncan.macrae@wolterskluwer.com; <sup>2</sup>Cactus Communications, Mumbai, India

**Conflict of Interest Disclosures** None reported.

#### Mitigating Subjectivity in Peer Review via Artificial Intelligence

Henry Gouk,<sup>1</sup> Nihar Shah<sup>2</sup>

**Objective** Peer review incurs a problem of subjectivity (also called *commensuration bias*)<sup>1,2</sup>: different reviewers put differing emphasis on the various criteria when combining them to make an overall recommendation. Such subjectivity and resulting inconsistencies in the reviews are said to

**Table 51. Comparison of Initial Manuscript Rejections According to Use of a Precheck Tool Prior to Submission**

Precheck status	Submissions, No.	Initial rejections, No. (%)
No precheck	6062	2073 (34.2)
Basic precheck	1661	333 (20.1)
Premium precheck	181	13 (7.3)

contribute to arbitrariness of the peer review process.<sup>1,2</sup> In a top-tier venue's peer review process, 2 types of inconsistencies due to such subjectivity were quantified, and an artificial intelligence (AI)-based strategy<sup>3</sup> was deployed to mitigate such subjectivity.

**Design** This study was conducted in the final decision phase of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence 2022, which is a top scientific conference in artificial intelligence that reviews full papers, is a terminal publication venue, and is considered at par with journals. Reviewers were asked to evaluate papers on 8 criteria and also provide an overall recommendation. The issue of subjectivity pertains to different reviewers using different mappings from the criteria to their overall recommendations. Subjectivity may also lead to the following inconsistency. For any review,  $r$ , let  $c_r$  denote the vector of criteria scores and  $o_r$  denote the numeric overall recommendation. For any criteria score vectors  $c$  and  $c'$ ,  $c > c'$  if every entry of  $c$  is at least as large as the corresponding entry of  $c'$  and at least 1 entry of  $c$  is strictly larger than the corresponding entry of  $c'$ . A pair of reviews,  $r$  and  $s$ , are said to be inconsistent if  $c_r > c_s$  and  $o_r \leq o_s$  or if  $c_r = c_s$  and  $o_r \neq o_s$ . They are strongly inconsistent if  $c_r > c_s$  and  $o_r < o_s$ . An AI technique<sup>3</sup> was deployed to mitigate such subjectivity. From the reviews, the AI learned a single mapping from criteria scores to overall recommendations that was representative of the entire set of reviews. Then, it applied the learned mapping to the criteria scores in each review to obtain an updated overall recommendation associated with each review. It mitigated subjectivity by virtue of the fact that the updated overall recommendations for all reviews were obtained via the same mapping. The design of the AI also ensured that the updated overall scores did not have the aforementioned inconsistencies.

**Results** There were 177 reviews in which the difference between the original overall recommendation provided by the reviewer and the updated overall recommendation computed by the AI was 2 or greater (on a 10-point scale); the (equivalents of the) associate editors for these reviews were notified. The maximum such difference was 7. The number of inconsistencies in the reviews is shown in **Table 52**.

**Conclusions** A large number of inconsistencies were found across reviewers in how criteria are mapped to overall recommendations. An AI method<sup>3</sup> was successfully deployed to mitigate subjectivity in peer reviews.

**Table 52. Inconsistencies in the Reviews in the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence 2022 in the Set of All Reviews Originally Submitted by the Reviewers and in the Set of All (Updated) Reviews After Applying the Artificial Intelligence (AI) Technique<sup>3</sup>**

Measure	Original reviews	After applying AI technique <sup>3</sup>
Papers, No.	5127	5127
Reviews per paper, mean (SD)	4.02 (0.69)	4.02 (0.69)
Inconsistencies across all pairs of reviews, No./total No. (%)	8,404,439/73,035,572 (11.51)	0
Number of strong inconsistencies across all pairs of reviews, No./total No. (%)	2,845,870/73,035,572 (3.90)	0

## References

- Lee CJ. Commensuration bias in peer review. *Philos Sci*. 2015;82(5):1272-1283. doi:10.1086/683652
- Kerr S, Tolliver J, Petree D. Manuscript characteristics which influence acceptance for management and social science journals. *Acad Manage J*. 1977;20(1):132-141. doi:10.5465/255467
- Noothigattu R, Shah N, Procaccia A. Loss functions, axioms, and peer review. *J Artif Intelligence Res*. 2021;70:1481-1515. doi:10.1613/jair.1.12554

<sup>1</sup>University of Edinburgh, Edinburgh, Scotland; <sup>2</sup>Carnegie Mellon University, Pittsburgh, PA, USA, nihars@cs.cmu.edu

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported by the US National Science Foundation CAREER award 1942124, which supports research on the fundamentals of learning from people with applications to peer review.

## Authorship and Contributorship

### Analysis of Gender Representation, Authorship Inflation, and Institutional Affiliation in Abstract Acceptance, 2017-2021

Joseph S. Puthumana,<sup>1</sup> Iman F. Khan,<sup>1</sup> Rafael F. P. Tiongco,<sup>1</sup> Siam K. Rezwan,<sup>1</sup> Rena Atayeva,<sup>1</sup> Jeffry Nahmias,<sup>2</sup> Sarah Jung,<sup>3</sup> Carisa M. Cooney<sup>1</sup>

**Objective** The goals of this study were to characterize and examine associations of author number, author gender, and institutional affiliation with ratings and acceptances of all abstracts submitted to one surgical education conference.

**Design** All abstracts submitted between 2017 and 2021 to the annual meeting of the Association for Surgical Education were retrospectively reviewed. Abstract data included mean rater scores (1 indicates lowest and 7, highest), acceptance status, complete author lists, and institutional affiliations, data to which Association for Surgical Education abstract reviewers were blinded. Last author affiliation was cross-referenced with top 40 National Institutes of Health (NIH)-funded institutions,<sup>1</sup> and Gender API<sup>2</sup> was used to code first and last author genders. One-way analysis of variance was

used to analyze continuous variables, Pearson  $\chi^2$  to analyze categorical variables, z score population proportions to compare submission-to-podium acceptance rates, and Pearson correlation coefficient to assess associations between abstract score and number of authors/institutions.

**Results** A total of 1162 abstracts were analyzed. A higher but statistically not significant number of authors per submission was found (4.90 in 2017 to 5.33 in 2021;  $P = .06$ ), and there was no difference in the number of institutions per submission (1.75 in 2017 to 1.83 in 2021;  $P = .98$ ) over time. Higher reviewer scores demonstrated positive but weak correlations with more authors ( $r_{1160} = 0.191$ ;  $P < .001$ ) and institutions ( $r_{1160} = 0.182$ ;  $P < .001$ ). Significantly higher scores were noted for abstracts with last authors affiliated with top 40 NIH-funded institutions (4.18 [0.96%] vs 3.72 [1.12%];  $P < .001$ ); these abstracts were accepted for podium presentation at a higher rate (42.1% vs 29.7%;  $P < .001$ ). Abstracts submitted by last authors affiliated with top 40 NIH-funded institutions had significantly more authors (5.46 [2.56%] vs 4.88 [2.42%];  $P < .001$ ). Women were first authors in 51.8% ( $n = 602$ ) and last authors in 35.4% ( $n = 411$ ) of all abstracts. Abstracts were rated significantly higher with women rather than men as first authors (3.98 [0.99%] vs 3.82 [1.12%];  $P = .01$ ) or last authors (4.01 [1.04%] vs 3.82 [1.10%];  $P = .005$ ) (Table 53). Across all years, abstracts submitted by women first or last authors were accepted more often as podium or plenary presentations.

**Conclusions** Abstracts whose last author had a top 40 NIH-funded institutional affiliation received significantly higher scores, possibly indicating increased tangible or intangible resources contributing to research efforts. This study demonstrated a nonsignificant trend toward more authors over the 5 years studied. Abstracts with women first and last authors were scored higher and were more frequently invited for plenary and podium presentations. While women composed the majority of first authors, women representation as last authors has not yet reached parity but increased over the course of the study despite the COVID-19 pandemic.

## References

1. NIH awards by location & organization. National Institutes of Health. <https://report.nih.gov/award/index.cfm>
2. Gender API. <https://gender-api.com/>

<sup>1</sup>Department of Plastic and Reconstructive Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA, ccooney3@jhmi.edu; <sup>2</sup>Department of Surgery, University of California, Irvine, Orange, CA, USA; <sup>3</sup>Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

**Conflict of Interest Disclosures** None reported.

## Numbers and Trends in Authorship of Published Meta-analyses, 1990-2019

Marios Papadakis<sup>1</sup>

**Objective** The aim of this study was to investigate the evolution of authorship trends in MEDLINE-indexed

**Table 53. Number of Women First and Last Authors by Year With Corresponding Podium/Plenary Presentation Acceptance Rates**

Factor	No. (%)				
	2017	2018	2019	2020	2021
Total submissions, No.	166	227	254	303	212
Total posters	80 (48.2)	95 (41.9)	101 (39.8)	101 (33.3)	101 (47.6)
Total podiums	43 (25.9)	50 (22.0)	50 (19.7)	69 (22.8)	100 (47.2)
Women first author submissions	88 (57.1)	109 (49.3)	115 (47.1)	166 (57.4)	124 (61.7)
Women first author podium acceptances	44 (50.0)	53 (48.6)	50 (43.5)	62 (37.3)	70 (56.5)
Men first author submissions	66 (42.9)	112 (50.7)	129 (52.9)	123 (42.6)	77 (38.3)
Men first author podium acceptances	33 (50.0)	40 (35.7)	49 (38.0)	37 (30.1)	29 (37.7)
Women last author submissions	53 (31.9)	73 (32.2)	92 (36.4)	107 (36.1)	85 (40.3)
Women last author podium acceptances	26 (49.1)	35 (47.9)	39 (42.4)	38 (35.5)	45 (52.9)
Men last author submissions	113 (68.1)	154 (67.8)	161 (63.6)	189 (63.9)	126 (59.7)
Men last author podium acceptances	54 (47.8)	60 (39.0)	62 (38.5)	62 (32.8)	56 (44.4)

meta-analyses over time. Although similar studies do exist,<sup>1</sup> they all have major limitations (cross-sectional nature, small sample sizes, etc) that do not allow for reliable conclusions to be drawn. This may be the first work to study authorship patterns in all MEDLINE-indexed meta-analyses published.

**Design** In PubMed, the search filters of study type and publication date were applied, and the search was restricted to meta-analyses published until December 31, 2019. Single research group names included in the article's title were considered equivalent to 1 coauthor. Articles without author names were excluded from the study.

**Results** A total of 116,710 meta-analyses were analyzed. The most meta-analyses per year were published in 2019 (1.2%), followed by 2018 (1.1%). The overall mean (SD) number of authors was 5.4 (4). The most common number of authors was 4, found in 16.4% of articles, followed by 5 authors (15.4%). The mean number of authors per article increased significantly over time, from 3 in 1990-1994 to 5.8 in 2015-2019. Single-author articles represented 24% of all articles in 1990-1994 and only 1.2% in 2015-2019. The number of articles authored by 15 or more authors increased from 0% in 1990-1994 to 3.2% in 2015-2019.

**Conclusions** The reported trend of authorship proliferation<sup>1,2</sup> was also observed in meta-analyses, with a current mean number of 5.8 authors per article. It is doubtful whether this increase can only be attributed to increasing research complexity.<sup>2</sup> Scientists should adhere to the existing

guidelines and include in the author list only contributors who qualify for inclusion. Journals should adopt more strict policies to confirm that only substantial contributions are getting credited with authorship.

## References

1. Gülen S, Fonnes S, Andresen K, Rosenberg J. Increasing number of authors in Cochrane reviews. *J Evid Based Med.* 2020;13(1):34-41. doi:10.1111/jebm.12371
2. Papatheodorou SI, Trikalinos TA, Ioannidis JPA. Inflated numbers of authors over time have not been just due to increasing research complexity. *J Clin Epidemiol.* 2008;61(6):546-551. doi:10.1016/j.jclinepi.2007.07.017

<sup>1</sup>Department of Surgery II, University Witten-Herdecke, Witten, Germany, marios\_papadakis@yahoo.gr

**Conflict of Interest Disclosures** None reported.

## Bias

### Development of the Quality Assessment of Prognostic Accuracy Studies (QUAPAS) Tool for Assessing Risk of Bias in Prognostic Accuracy Studies

Jenny Lee,<sup>1</sup> Frits Mulder,<sup>2</sup> Mariska Leeflang,<sup>1</sup> Robert Wolff,<sup>3</sup> Penny Whiting,<sup>4</sup> Patrick M. Bossuyt<sup>1</sup>

**Objective** Systematic reviews of prognostic accuracy studies often use the QUADAS-2 (Quality Assessment of Diagnostic

Accuracy Studies) tool<sup>1</sup> to assess risk of bias and applicability of included studies because no comparable instrument exists for prognostic accuracy studies. The QUAPAS (Quality Assessment of Prognostic Accuracy Studies) tool is an adaptation of QUADAS-2 for prognostic accuracy studies.

**Design** Six experienced reviewers and methodologists in the area of test evaluation and/or development of risk of bias tools participated in the development of QUAPAS. They used risk of bias tools suggested for systematic reviews by the Cochrane Prognosis Methods Group and Diagnostic Test Accuracy Working Group. Using QUADAS-2 as a starting point, domains and signaling questions from QUIPS (Quality in Prognosis Studies)<sup>2</sup> and PROBAST (Prediction Model Risk of Bias Assessment Tool)<sup>3</sup> were evaluated in parallel to collate a unique list of signaling questions for each domain. Additional signaling questions based on relevant sources of bias were included. The steering group of 6 experts conducted and reviewed 3 rounds of modifications before arriving at the final set of domains and signaling questions. The authors further shared QUAPAS with and invited written feedback from 10 researchers and potential end users.

**Results** QUAPAS was developed to be used as QUADAS-2: specify the review question, tailor the tool, draw a flow diagram, judge risk of bias, and identify applicability concerns. Risk of bias was judged across 5 domains: participants, index test, outcome, flow and timing, and analysis (**Table 54**). Signaling questions assisted the final judgment for each domain. Applicability concerns were assessed for the first 4 domains. QUAPAS was used in parallel

**Table 54. The QUAPAS Tool**

Domain	Participants	Index test	Outcome	Flow and timing	Analysis
Description	Describe methods for recruiting participants  Describe participants (previous testing, presentation, intended use of index test, and setting)	Describe the index test (definition, context of use, method of measurement and interpretation)	Describe the outcome (definition, method of measurement and interpretation)	Describe any participants lost to follow-up or excluded from the analysis  Describe the time horizon from the index test to the outcome	Describe the statistical methods
Signaling questions (yes, no, unclear)	Was a consecutive or random sample of participants enrolled?  Was a case-control design avoided?  Did the study avoid inappropriate selection criteria?	Was the method used to perform the index test valid and reliable?  Was the method for performing the index test the same for all participants?  Were the index test results interpreted without knowledge of the outcome?  If a threshold was used, was it prespecified?	Was the method used to measure the outcome valid and reliable?  Was the method for measuring the outcome the same for all participants?  Was the outcome measured without knowledge of the index test results?	Did all participants receive the index test?  Was treatment avoided after the index test was measured?  Was the time horizon sufficient to capture the outcome?  Was information on the outcome available for all participants?	Were all enrolled participants included in the analysis?  If data were missing, were appropriate methods used?  Were appropriate methods used to account for censoring?  In case of competing events, were appropriate methods used to account for them?
Risk of bias (high, low, unclear)	Could the selection of participants have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could measurement of the outcome have introduced bias?	Could the study flow have introduced bias?	Could the analysis have introduced bias?
Concerns about applicability (high, low, unclear)	Are there concerns that the participants do not match the review question?	Are there concerns that the index test, its conduct, interpretation, or threshold differ from the review question?	Are there concerns that the outcome does not match the review question?	Are there concerns that the time horizon does not match the review question?	

Abbreviation: QUAPAS, Quality Assessment of Prognostic Accuracy Studies.

with QUADAS-2 and QUIPS in a systematic review evaluating the accuracy of a noninvasive liver test in prognosis of fibrosis and mortality. Use of QUAPAS improved the risk assessment of the flow and timing domain by the addition of a new signaling question and flagged studies at risk of bias in the new analysis domain. Judgment of risk of bias in the analysis domain was found challenging owing to sparse reporting of statistical methods.

**Conclusions** QUAPAS may provide future systematic reviewers and readers with a reliable tool that can assess bias and applicability concerns in prognostic accuracy studies. Use of a systematically tailored tool will hopefully improve the quality assessment process and help produce more robust evidence base for prognostic tests.

## References

1. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-20110180-00009
2. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med.* 2013;158(4):280-286. doi:10.7326/0003-4819-158-4-201302190-00009
3. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170(1):51-58. doi:10.7326/M18-1376

<sup>1</sup>Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, the Netherlands, j.a.lee@amsterdamumc.nl;

<sup>2</sup>Department of Vascular Medicine, Cardiovascular Sciences, Amsterdam, the Netherlands; <sup>3</sup>Kleijnen Systematic Reviews Ltd, Escrick, UK; <sup>4</sup>Bristol Medical School, University of Bristol, Bristol, UK

**Conflict of Interest Disclosures** Patrick M. Bossuyt is an advisory board member of the Peer Review Congress but was not involved in the review or decision of this abstract. No other disclosures were reported.

**Funding/Support** The LITMUS project received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement 777377. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA.

**Role of the Funder/Sponsor** The funders had no role in the study.

**Acknowledgment** The authors thank Jill Hayden, PhD (Dalhousie Medical School), for her valuable contributions to the earlier stages of developing QUAPAS. They also thank the volunteers for reviewing the tool.

## Citations

### Citations of Human Gene Research Articles That Describe Wrongly Identified Nucleotide Sequences

Yasunori Park,<sup>1</sup> Jennifer Anne Byrne<sup>1,2</sup>

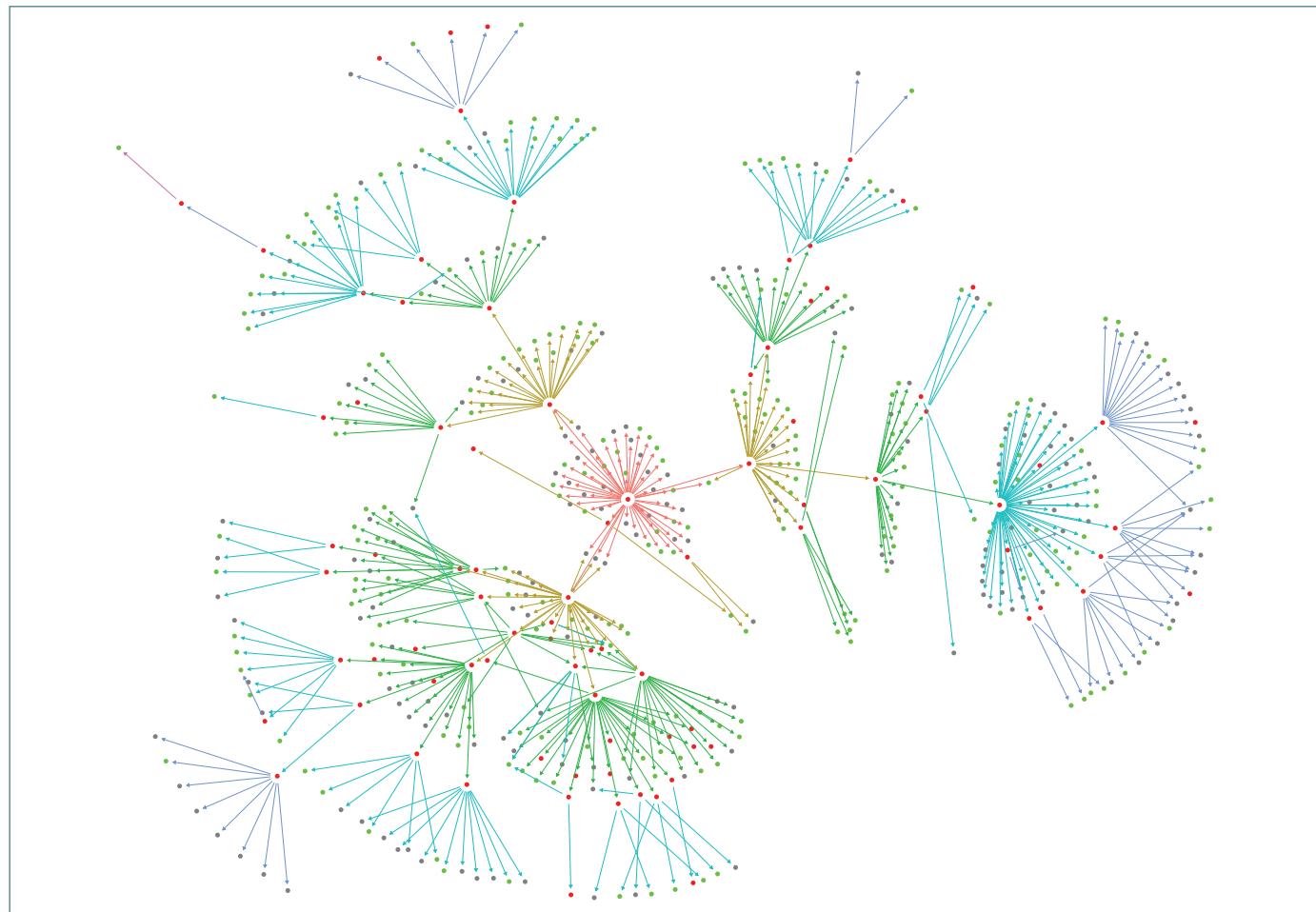
**Objective** Preclinical human gene research articles that describe wrongly identified nucleotide sequence reagents provide incorrect information, and these research articles could be manufactured by paper mills that "have been alleged to mass-produce fraudulent manuscripts for publication."<sup>1(p2)</sup> Such problematic articles can be highly cited<sup>1</sup>; however, the features and consequences of these citations are largely unknown. The authors investigated this question by analyzing citations of problematic human gene research articles and of literature reviews that cite these articles.

**Design** A human gene research article cited in PubMed (ie, PMID 25721211) with wrongly identified nucleotide sequences was selected as an index case<sup>2</sup> to build citation networks in R-studio using Google Scholar citations prior to March 31, 2022. The citing articles were screened for wrongly identified sequences<sup>1</sup> and citation context.<sup>3</sup> Problematic articles with wrongly identified sequences were examined through up to 6 citation generations. Because 95 literature reviews cited problematic articles, the authors also analyzed 13 literature reviews that focused on human genes. All review references were screened for wrongly identified nucleotide sequences<sup>1</sup> to identify problematic references. Each review text was examined to identify statements that cited problematic references. Publications that cited each literature review were examined to identify review citations that reflected information from problematic references.

**Results** After analysis of the citations in PMID 25721211 through up to 6 citation generations, 87 cited problematic articles (**Figure 23**) published in 50 journals were identified. As previously reported,<sup>1</sup> most problematic articles (79 of 87 [91%]) were authored by teams from hospitals in China. Ninety-three citations of problematic articles by other problematic articles were identified. A total of 360 statements were identified in 338 citing articles that were supported by problematic articles, typically in the Discussion section (183 of 360 [51%]) or Introduction section (133 of 360 [41%]). The 13 human gene literature reviews cited a total of 1887 references that included 206 problematic articles (11%). Between 1 and 13 claims per review (82 claims in total) were supported by problematic references. The 13 reviews have been cited 1843 times, in which 3 citations reflected claims from 3 problematic references. The 206 problematic references have been cited 31,914 times, including by 5 clinical trial articles. Problematic references were also cited by 78 patent families and 9 Wikipedia entries.

**Conclusions** After analysis of the citation network for 1 problematic gene research article, 87 problematic articles and 93 citations of problematic articles by other problematic articles were identified (**Figure 23**). It was further demonstrated that 13 literature reviews of human genes referenced 206 problematic articles that were in turn cited 31,914 times. Although infrequent, subsequent literature review citations can reflect information from problematic review references.

**Figure 23. Citation Network for Article Cited in PubMed**



The article cited in PubMed (ie, PMID 25721211) is located in the center. Publications are represented by circles colored according to the presence (red) or absence (green) of 1 or more wrongly identified nucleotide sequences or the absence of any nucleotide sequences (gray). Arrows indicate citations of problematic articles with wrongly identified nucleotide sequences and extend from cited to citing articles. Citations of PMID 25721211 (first-generation citations) are shown by red arrows, whereas second-, third-, fourth-, fifth-, and sixth-generation citations are shown by olive, green, aqua, blue, and pink arrows, respectively. Some problematic articles are yet to be cited.

## References

1. Park Y, West RA, Pathmendra P, et al. Identification of human gene research articles with wrongly identified nucleotide sequences. *Life Sci Alliance*. 2022;5:e202101203. doi:10.26508/lsa.202101203
2. Kempen JH. Appropriate use and reporting of uncontrolled case series in the medical literature. *Am J Ophthalmol*. 2011;151:7-10. doi:10.1016/j.ajo.2010.08.047
3. Zhang G, Ding Y, Milojević S. Citation content analysis (CCA): a framework for syntactic and semantic analysis of citation content. *J Am Soc Inf Sci Tec*. 2013;64:1490-1503. doi:10.1002/asi.22850

<sup>1</sup>Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia; <sup>2</sup>NSW Health Statewide Biobank, NSW Health Pathology, Camperdown, New South Wales, Australia, jennifer.byrne@health.nsw.gov.au

## Conflict of Interest Disclosures

**Funding/Support** Jennifer Anne Byrne acknowledges grant funding from the National Health and Medical Research Council of Australia, Ideas grant ID APP1184263.

**Role of the Funder/Sponsor** The funder played no role in the study design, data collection, management, analysis, or interpretation and will play no role in the writing of any report or the decision to submit the report for publication.

## Conflict of Interest

### Association Between Commercial Funding and Estimated Intervention Effects in Randomized Trials: The COMFIT Study

Camilla H. Nejstgaard,<sup>1,2</sup> Gemma Clayton,<sup>3</sup> Andreas Lundh,<sup>1,2,4</sup> Iosief Abraha,<sup>5</sup> Susan Armijo-Olivo,<sup>6,7</sup> Isabelle Boutron,<sup>8,9</sup> Robin Christensen,<sup>10,11</sup> Bruno R. da Costa,<sup>12,13,14</sup> Greta G. Cummings,<sup>15</sup> Agnes Dechartres,<sup>16</sup> Carlos Flores-Mir,<sup>17</sup> Anders D. Frost,<sup>1,2</sup> Toshi A. Furukawa,<sup>18</sup> Robin Haring,<sup>19,20</sup> Lisa Hartling,<sup>21</sup> John P. A. Ioannidis,<sup>22,23</sup> Mihaela Ivosevic,<sup>1,2</sup> Perrine Janiaud,<sup>23,24</sup> David R. T. Laursen,<sup>1,2</sup> Helene Moustgaard,<sup>1,2</sup> M. Hassan Murad,<sup>25</sup> Matthew J. Page,<sup>19</sup> Philippe Ravaud,<sup>8,9</sup> Humam Saltaji,<sup>17</sup> Jelena Savović,<sup>3,26</sup> Yasushi Tsujimoto,<sup>18,27</sup> Zhen Wang,<sup>25</sup> Asbjørn Hróbjartsson<sup>1,2</sup>

**Objective** Commercial companies often fund randomized clinical trials, and methodological studies suggest that this funding may influence trial results and conclusions.<sup>1</sup> However, such methodological studies are often at risk of confounding given that they compare trials that differ for reasons other than funding source. The risk of confounding can be minimized in meta-epidemiological studies that compare similar trials within meta-analyses. The Commercial Funding in Trials (COMFIT) study was initiated based on a consortium of researchers who shared data sets with information on meta-analyses and trials included in meta-epidemiological studies (details available in the protocol<sup>2</sup>). The primary aim of COMFIT was to investigate the impact of commercial funding on estimated intervention effects in randomized clinical trials.

**Design** The study approach was to identify meta-epidemiological studies and combine and reanalyze their meta-analysis and trial data. Studies with data on trial funding source and results were included (eg, primarily by investigating funding source or investigating a different trial characteristic and adjusting for funding source). Five bibliographic databases and other sources (eg, conference proceedings) were searched. Meta-epidemiological study authors were contacted and invited to join the COMFIT consortium and share their unpublished data. Construction of the COMFIT database involved, for example, merging data from included studies and checking data quality. Noninformative meta-analyses (eg, meta-analyses in which all trials had the same funding type) and overlapping or correlated meta-analyses and trials (eg, multiple meta-analyses included from the same systematic review) were removed. Bayesian hierarchical models were used to estimate the ratio of odds ratios (RORs) and heterogeneity statistics (to describe within- and between-study heterogeneity). A ROR <1 indicated exaggerated intervention effects in trials with commercial funding. Unadjusted and adjusted analyses using bias domains (low vs high risk of bias from sequence generation, allocation concealment, blinding, incomplete outcome data, selective reporting, and overall risk of bias) and trial sample size were performed. Subgroup and sensitivity analyses (eg, using the Sterne 2-step approach<sup>3</sup>) were performed.

**Results** The COMFIT database contained data on 554 meta-analyses, with 4936 trials from 17 meta-epidemiological studies. Meta-analyses included a median (range) of 7 (2–68) trials. In total, 1368 trials (28%) had noncommercial funding, 1840 trials (37%) had sole commercial or mixed commercial and noncommercial funding, and the remaining 1728 trials (35%) did not have funding sources reported. Using data from the largest meta-epidemiological study (20% of meta-analyses, 19% of trials), the ROR for the association between commercial funding and estimated intervention effects was 0.90 (95% CI, 0.84–0.96) (preliminary findings, not the primary analysis).

**Conclusions** The COMFIT database may enable comprehensive analyses of the impact of commercial funding

on trial results. The primary analyses will have a markedly increased statistical power and reduced risk of confounding and reporting bias compared with previous studies.

## References

1. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev*. 2017;2:MR000033. doi:10.1002/14651858.MR000033.pub3
2. Nejstgaard CH, Lundh A, Abdi S, et al. Combining meta-epidemiological study datasets on commercial funding of randomised clinical trials: database, methods, and descriptive results of the COMFIT study. *Res Synth Methods*. 2022;13(2):214–228. doi:10.1002/jrsm.1527
3. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in “meta-epidemiological” research. *Stat Med*. 2002;21:1513–1524.

<sup>1</sup>Centre for Evidence-Based Medicine Odense and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark, cahansen@health.sdu.dk; <sup>2</sup>Open Patient Data Exploratory Network, Odense University Hospital, Odense, Denmark; <sup>3</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; <sup>4</sup>Department of Infectious Diseases, Hvidovre Hospital, Hvidovre, Denmark; <sup>5</sup>Servizio Immunotrasfusionale, USL Umbria 2, Foligno, Italy; <sup>6</sup>University of Applied Sciences, Faculty of Business and Social Sciences, Hamburg, Germany; <sup>7</sup>Faculty of Medicine and Dentistry and Faculty of Rehabilitation Medicine, Department of Physical Therapy, Rehabilitation Research Center, Edmonton, AB, Canada; <sup>8</sup>Université de Paris, INSERM, INRAE, CNAM, Centre of Research in Epidemiology and Statistics, Paris, France; <sup>9</sup>Cochrane France, Paris, France; <sup>10</sup>Section for Biostatistics and Evidence-Based Research, Parker Institute, Frederiksberg and Bispebjerg Hospital, Frederiksberg, Denmark; <sup>11</sup>Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark; <sup>12</sup>Applied Health Research Centre, Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, ON, Canada; <sup>13</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; <sup>14</sup>Institute of Primary Health Care, University of Bern, Bern, Switzerland; <sup>15</sup>Faculty of Nursing, University of Alberta, Edmonton, AB, Canada; <sup>16</sup>Sorbonne Université, INSERM, Institut Pierre Louis d’Epidémiologie et de Santé Publique, APHP, Sorbonne Université, Hôpital Pitié Salpêtrière, Département de Santé Publique, Centre de Pharmacoépidémiologie, Paris, France; <sup>17</sup>Department of Dentistry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada; <sup>18</sup>Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; <sup>19</sup>School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia; <sup>20</sup>European University of Applied Sciences, Brühl, Germany; <sup>21</sup>Alberta Research Centre for Health Evidence, Department of Pediatrics, University of Alberta, Edmonton, AB, Canada; <sup>22</sup>Stanford Prevention Research Center Department of Medicine, Stanford University, Palo Alto, CA, USA; <sup>23</sup>Meta-Research Innovation Center at Stanford, Stanford University, Palo Alto, CA, USA; <sup>24</sup>Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland; <sup>25</sup>Evidence-Based Practice Center, Mayo Clinic, Rochester, MN, USA; <sup>26</sup>NIHR Applied Research Collaboration West, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK; <sup>27</sup>Scientific Research Works Peer Support Group, Japan

**Conflict of Interest Disclosures** Asbjørn Hróbjartsson and Andreas Lundh are currently members of the TACIT steering group to develop a tool for addressing conflicts of interest in trials. Toshi A. Furukawa reported receiving grants and personal fees from Mitsubishi-Tanabe and Shionogi, receiving personal fees from Sony outside the submitted work, and having patent 2020-548587 concerning smartphone cognitive behavioral applications pending and intellectual properties for the Kokoro application licensed to Mitsubishi-Tanabe. Jelena Savović was paid by Core Models to deliver an online teaching session of basic evidence synthesis methods. John P. A. Ioannidis is co-director and Isabelle Boutron is an advisory board member of the Peer Review Congress, but they were not involved in the review or decision for this abstract.

## Evaluation of Journal Editor Conflict of Interest Disclosures and Remuneration Transparency in Oncology and Cardiology

Paul J. Hauptman,<sup>1</sup> Chelsea Price,<sup>1</sup> Eric Heidel<sup>1</sup>

**Objective** Recent data suggest that authors may not be transparent with reporting financial conflicts of interest.<sup>1,2</sup> However, there is a paucity of data about journal editors,<sup>2</sup> who are the ultimate decision-makers regarding publication. This topic is particularly important among specialties with significant academic-industry interactions.

**Design** A cross-section of the top journals in oncology (100) and cardiology (100) were selected based on their journal impact factors (JIFs).<sup>3</sup> The JIF and total number of citations for each journal were collated and the primary editors (editor in chief, executive editor, or similar designation) for each journal were identified through journal websites; for journals listing more than a single primary editor (14), each editor was considered separately. Every editor (216) was queried on the Center for Medicare and Medicaid Services Open Payments Database (<https://openpaymentsdata.cms.gov>) for the most recently reported total payments, general payments, and research payments or associated research funding since 2014. Exclusively international editors (33 for oncology and 45 for cardiology) were not included in analyses. One journal was removed from the oncology data set because of a suspected error in JIF calculation. Payments for each specialty were analyzed using medians and IQRs. Journal websites were examined for editors with more than US \$5000 in total payments to assess for transparency.

**Results** Of 78 editors in oncology, 40 accepted general payments and 24 collected research payments (18 received both); of 60 editors for cardiology, 42 received general payments and 18 received research payments (12 received both). Median general payments were similar between the 2 specialties (Table 55), but research payments were higher for oncology. In journals with editors receiving payments of any type greater than \$5000 (54), generic conflict of interest policy statements appeared on 25 journal websites; 3 listed specific dollar amounts or specific companies.

**Conclusions** Despite increased focus on potential conflicts of interest in journal editing, a significant proportion of editors of top-tier oncology and cardiology journals receive payments from industry, and the nature of these relationships

**Table 55. General and Research Payments to Editors by Specialty**

Specialty	Payment type (No. of editors)	Payment amount, median (IQR), US \$
Oncology	Total (78)	80 (0-26,473)
	General (40)	2836 (96-26,401)
	Research (24)	84,477 (12,577-297,124)
Cardiology	Total (60)	1750 (19-22,191)
	General (42)	3096 (193-22,191)
	Research (18)	32,196 (4450-100,287)

are not transparent to authors or the readership. Given the role that editors play in the publication process, specific conflicts and recusal policies should be more clearly delineated.

## References

1. Baraldi JH, Picozzo SA, Arnold JC, Volarich K, Giofriddo MR, Piper BJ. A cross-sectional examination of conflict-of-interest disclosures of physician-authors publishing in high-impact US medical journals. *medRxiv*. Preprint posted online January 1, 2022. doi:10.1101/2021.09.12.21263468
2. Dal-Ré R, Caplan AL, Marusic A. Editors' and authors' individual conflicts of interest disclosure and journal transparency: a cross-sectional study of high-impact medical specialty journals. *BMJ Open*. 2019;9:e029796. doi:10.1136/bmjopen-2019-029796
3. Clarivate Analytics. 2020 Journal Impact Factor. *Journal Citation Reports*. Clarivate; 2021.

<sup>1</sup>University of Tennessee Graduate School of Medicine, Knoxville, TN, USA, phauptman@utmck.edu

**Conflict of Interest Disclosures** Paul J. Hauptman reported receiving remuneration from Corvia Medical (DSMB), scPharma and LivaNova (clinical events committees), and Array Biopharma (steering committee) for work related to clinical trials as indicated. No other disclosures were reported.

## Data Presentation and Graphical Display

### Editors' Perspectives on Adding a Results Table and Limitations Section to Medical Journal Abstracts: A Qualitative Study

Steven Woloshin,<sup>1,2</sup> Rebecca J. Williams,<sup>2</sup> Lisa Bero<sup>2,3</sup>

**Objective** To assess editors' experience with and openness to refining journal article abstracts by adding a results table and a limitations header to improve abstract readability and informational content.<sup>1-3</sup>

**Design** General medical journals were selected based on journal impact factor rankings: all top 10 ranked journals and 5 among those ranked in the third quartile, published in English, with multiple issues per year, and using a structured abstract. Only one journal from the JAMA Network was selected. Semistructured interviews were conducted with the

editor recommended by the journal's editor in chief. The study ethics approval (exemption) was provided by Dartmouth.

**Results** Eleven of the 15 invited journals participated (9 from the top 10 and 2 of 5 from the third quartile by impact factor). Interviews were conducted with 4 editors in chief, 3 executive editors, and 4 other editor types by S.W. on a web conferencing platform (1 editor responded in writing) from February 4 to March 4, 2022, and lasted 15 to 20 minutes. Calls were recorded and autotranscribed. All study authors reviewed the full interview transcripts, R.J.W. summarized key themes from transcripts, and all authors reached consensus on abstract results tables key themes (**Table 56**). One journal had experience publishing abstract results tables. There was strong interest in a limitations header and few concerns about it having any potential harms.

**Conclusions** These findings provide preliminary support for a trial evaluating the addition of results tables and limitations to abstracts. Limitations of this study are that it may not be representative of all journals and interviews did not include abstract readers or authors.

## References

1. Bauchner H, Henry R, Golub RM. The restructuring of structured abstracts: adding a table in the results section. *JAMA*. 2013;309(5):491-492. doi:10.1001/jama.2013.76
2. Ertl N, Gazette K. A new way of documenting scientific data from medical publications. *Karger Gaz*. 1969;20(27):1-3.
3. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited: a progress report. *Ann Intern Med*. 1990;113(1):69-76. doi:10.7326/0003-4819-113-1-69

<sup>1</sup>Center for Medicine and the Media, The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at

**Table 56. Key Themes From Interviews With Editors of Medical Journals on Adding Structured Results Tables to Journal Article Abstracts**

Category	Common Themes Expressed
Potential benefits	<ul style="list-style-type: none"><li>• Scientific communication: general enthusiasm for results table to improve communication of primary outcome results and limiting selective outcome reporting</li><li>• Reader engagement: may entice readers to read more details about the methods and results in the full article</li></ul>
Potential harms	<ul style="list-style-type: none"><li>• Burden: may increase workload on both editorial staff and authors</li><li>• Scientific concerns: table may not accommodate all types of study designs and methods; care needed to avoid selective results reporting and spin</li><li>• Reader engagement: may discourage readers from accessing the full article if key content is already in abstract</li></ul>
Barriers to implementation	<ul style="list-style-type: none"><li>• Tables may take up too much space</li><li>• Formatting requirements: PubMed and some publishers do not allow tables in the abstract, and it is not desirable to have 2 versions of abstract</li><li>• Author and editor time constraints</li></ul>
Other/general	<ul style="list-style-type: none"><li>• Strong interest in participating in trial of results tables</li><li>• Several journals noted possible overlap with visual/graphic abstracts</li></ul>

Dartmouth, Lebanon, NH, USA, steven.woloshin@dartmouth.edu;  
<sup>2</sup>The Lisa Schwartz Foundation for Truth in Medicine; <sup>3</sup>Center for Bioethics and Humanities, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

**Conflict of Interest Disclosures** Steven Woloshin serves on the Cochrane Collaboration and the *JAMA Internal Medicine* editorial boards. Lisa Bero is Senior Research Integrity Editor, Cochrane, and serves on the Cochrane Editorial Board; an academic editor, Meta-Research, *PLoS Biology*; and a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. No other disclosures were reported.

**Acknowledgment** This paper is dedicated to the memory of Lisa Schwartz, MD, MS, our remarkable partner, colleague, and friend whose work inspired this project. The authors thank the journal editors and staff who participated in interviews.

## Data Sharing and Access

### Assessment of Time and Resources Required to Share Data for 2 Individual Participant Data Meta-analyses

Anna Lene Seidler,<sup>1</sup> Jonathan G. Williams,<sup>1</sup> Mason Aberoumand,<sup>1</sup> Kylie E. Hunter,<sup>1</sup> James Sotiropoulos,<sup>1</sup> Sol Libesman,<sup>1</sup> Angie Barba,<sup>1</sup> Angela C. Webster<sup>1</sup>

**Objective** To quantify the administrative demands and timeliness of sharing individual participant data and identify the main choke points for researchers and contracts departments.

**Design** This mixed-methods study conducted in 2022 included 2 case studies of individual participant data meta-analyses, Individual Participant Data on Cord Management at Preterm Birth (iCOMP) and Transforming Obesity Prevention for Children (TOPCHILD), for which data were requested from international trials from 2019 until May 2022. Each contact with trial investigators was logged in custom-built software. From these logs, contact time points, study characteristics (eg, year, sample size, country), email texts, and data-sharing status were extracted. Researchers responsible for negotiating agreements were interviewed to identify choke points. The median, IQR, and range of time from the first data request until data receipt were calculated, and the number of emails required was tallied. Emails, data-sharing agreements, reasons data could not be shared, and interview responses were analyzed qualitatively using a thematic analysis to identify choke points and facilitators.

**Results** For 71 of 72 included trials (50 iCOMP trials [69%], 22 TOPCHILD trials [31%]), data were directly requested and supplied from investigators. Only 1 data set was available as a publication supplement; none used data repositories. The median (IQR) time from requesting to receiving data was 88 days (130 days). The longest completed request took 831 days (ie, >2 years). A mean of 23 emails (SD, 25; range, 4-49 emails) were sent until data were received. While some data were shared in the requested format (26 trials [36%]), other data sets had to be recoded (46 trials [64%]), and substantial contact with the investigators was required for this. Workload

and duration to set up data sharing agreements were reduced if institutions had executed agreements together previously. For 3 studies, data were not shared despite investigators being eager to do so because institutional approval could not be obtained. For iCOMP, 7 requests remained in progress after 854 days; for TOPCHILD, 9 requests were in progress after 272 days. Most data sharing agreements were similar in content but varied in detail and length (3–36 pages). The largest choke points included differences in regulations across jurisdictions and varying stakeholder expectations of rights and responsibilities (eg, institutional expectations about data storage, authorship rights). However, most emails dealt with minor change requests and clarifications.

**Conclusions** This study's results demonstrate major delays in the process of data sharing from trials with requests remaining unresolved after 2 years. The similarities identified across agreements indicate opportunities for standardization. Data sharing delays consume financial resources and impair the timeliness of research, which is particularly problematic for research endeavors linked to emergency situations (such as the COVID-19 pandemic). Many researchers are willing, or required by journals and funders, to share their data. To overcome costly choke points and delays, streamlined standards and infrastructure are needed.

<sup>1</sup>NHMRC Clinical Trials Centre at the University of Sydney, Camperdown, New South Wales, Australia, lene.seidler@sydney.edu.au

**Conflict of Interest Disclosures** Anna Lene Seidler, Jonathan G. Williams, Mason Aberoumand, Kylie E. Hunter, James Sotiropoulos, Sol Libesman, and Angie Barba are affiliated with the iCOMP and TOPCHILD collaborations. Angela C. Webster is affiliated with the TOPCHILD collaboration. No other disclosures were reported.

**Funding/Support** Anna Lene Seidler receives research funding support from a National Health and Medical Research Council (NHMRC) investigator grant (GNT2009432). Kylie E. Hunter receives research funding support via 2 scholarships administered by the University of Sydney (Postgraduate Research Supplementary Scholarship in Methods Development [SC3504] and Research Training Program Stipend [SC3227]). TOPCHILD is supported by an NHMRC ideas grant (GNT1186363); iCOMP is supported by an NHMRC project grant (GNT1163585).

**Role of the Funder/Sponsor** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgment** We acknowledge all members of the iCOMP collaboration (<https://www.icompsstudy.org/>) and TOPCHILD collaboration (<https://www.opchildcollaboration.org/>) for their assistance with preparing and executing data sharing agreements and sharing their data despite the barriers they encountered.

## Bibliometric and Language Factors Associated With Studies With Authors Who Share Data Requested for a Systematic Review

Carolina M. Ferreira,<sup>1</sup> Natália D. Reis,<sup>1</sup> Marcus T. Silva,<sup>2</sup> Taís F. Galvão<sup>1</sup>

**Objective** Sharing data requested by contacting the corresponding author is a good practice in research integrity and highly expected by authors, but it can yield erratic results when used in systematic reviews.<sup>1</sup> Bibliometric factors are indirect measures of scientific report and research quality. The objective of this study was to assess whether bibliometric factors were associated with the availability and sharing of data requested for a previous systematic review.

**Design** This was an analysis of a cross-sectional study of studies included in a previous systematic review performed by the same author group in 2021, which was aimed to assess the prevalence of childhood obesity in Brazil.<sup>2</sup> Standard methods were followed for identifying and including studies, which were all assessed for methodological quality.<sup>2</sup> In cases where data were needed for the analysis, the study's authors were contacted and the outcome of the request was assessed.<sup>2,3</sup> This analysis included studies with authors who were contacted to request needed data from April 2018 to May 2020.<sup>3</sup> The total number of citations on Google Scholar, the Journal Citation Reports 2020 Impact Factor (JIF), and the language(s) that the main report was published in were collected to September 2021. The outcome was success in receiving the requested data, and the association with studies' bibliometric factors was investigated. The *t* test was used to assess the association of the outcome with the mean number of citations and JIF, and Pearson  $\chi^2$  or Fishers exact tests were used to test for language. We used Stata, version 14.2 (StataCorp LLC) for all calculations.

**Results** Of the 9394 retrieved records, 163 study authors were contacted and 51 studies' authors sent the requested data. The mean (SD) number of citations was higher in studies that failed to send data (50.5 [63.0]; 112 studies) than in studies that sent requested data (31.8 [37.0]; 51 studies), but this difference was not statistically different ( $P = .05$ ). For 97 studies published in journals with a JIF, no difference was observed according to success (2.3 [1.1]; 31 studies) or failure (2.2 [1.3]; 66 studies) ( $P = .69$ ). Success was significantly lower in studies published in Portuguese (28 of 109 studies [25.7%]) than in other languages (23 of 54 studies [42.6%]) ( $P = .03$ ), and no difference in success was observed for studies published in English (33 of 94 studies [35.1%] vs 18 of 69 studies [26.1%];  $P = .22$ ) and Spanish (1 of 7 studies [14.3%] vs 50 of 156 studies [32.1%];  $P = .44$ ) in comparison with other languages (**Table 57**).

**Conclusions** Despite representing quality of scientific reporting, bibliometric factors such as number of citations and JIF were not associated with the success of obtaining data through contacting the author. Studies published in Portuguese were less likely to share the requested data, possibly reflecting less involvement in research dissemination. Availability of full data sets should be encouraged for all scientific publications to improve value and reduce waste on research.

## References

1. Meursinge Reynders R, Ladu L, Di Girolamo N. Contacting of authors modified crucial outcomes of systematic reviews

**Table 57. Success and Failure in Receiving Data According to Studies' Bibliometric Factors**

Bibliometric factors	No.	Success, No. (%) (n = 51)	Failure, No (%) (n = 112)	P value
Google Scholar citations, mean (SD), No.	163	31.8 (37.0)	50.5 (63.0)	.05 <sup>a</sup>
Impact factor, mean (SD) <sup>b</sup>	97	2.3 (1.1)	2.2 (1.3)	.69 <sup>a</sup>
Available in Portuguese <sup>c</sup>	109	28 (25.7)	81 (74.3)	.03 <sup>d</sup>
Not available in Portuguese <sup>c</sup>	54	23 (42.6)	31 (57.4)	
Available in English <sup>c</sup>	94	33 (35.1)	61 (64.9)	.22 <sup>d</sup>
Not available in English <sup>c</sup>	69	18 (26.1)	51 (73.9)	
Available in Spanish <sup>c</sup>	7	1 (14.3)	6 (85.7)	.44 <sup>e</sup>
Not available in Spanish <sup>c</sup>	156	50 (32.1)	106 (68.0)	

<sup>a</sup>Two-sample t test.

<sup>b</sup>Of the 97 studies published in journals with an impact factor, 31 studies successfully responded to data requests and 66 failed to send data.

<sup>c</sup>Papers could be available in more than 1 language.

<sup>d</sup>Pearson χ<sup>2</sup> test.

<sup>e</sup>Fisher exact test.

but was poorly reported, not systematic, and produced conflicting results. *J Clin Epidemiol.* 2019;115:64-76.

2. Ferreira CM, Reis NDD, Castro AO, et al. Prevalence of childhood obesity in Brazil: systematic review and meta-analysis. *J Pediatr (Rio J).* 2021;97(5):490-499.

3. Dutra dos Reis N, Müller Ferreira C, Silva MT, Galvão TF. Frequency of receiving requested data for a systematic review and associated factors: a cross-sectional study. *Account Res.* 2022;29(3):165-177.

<sup>1</sup>School of Pharmaceutical Sciences, State University of Campinas, Campinas, São Paulo, Brazil, taisgalvao@gmail.com; <sup>2</sup>Postgraduate Program of Pharmaceutical Sciences, University of Sorocaba, Sorocaba, São Paulo, Brazil

**Conflict of Interest Disclosures** None reported.

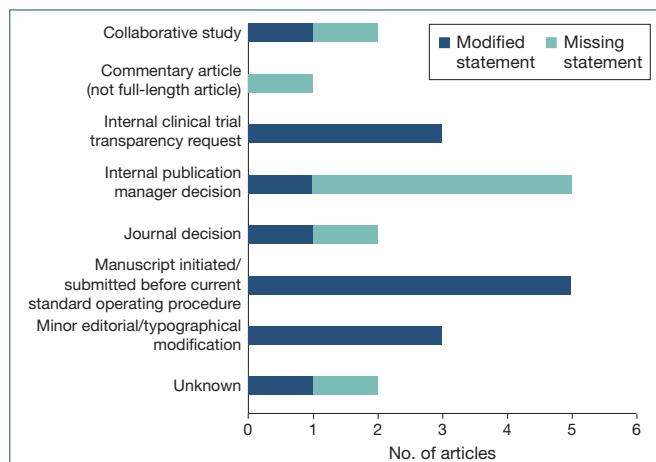
**Funding/Support** This work was supported by the National Council for Scientific and Technological Development (CNPq) under grant 4408652017-4. Taís F. Galvão receives productivity scholarship from CNPq (grant 3064482017-7).

### Data Sharing Statement Modifications in Manuscripts Reporting Interventional Clinical Trials Sponsored by a Global Biopharmaceutical Company

Colin McKinnon,<sup>1</sup> Jesse Potash,<sup>2</sup> Callan Fromm,<sup>2</sup> Teodor G. Paunescu,<sup>3</sup> Hajin Yang,<sup>2</sup> Ingeborg Cil,<sup>4</sup> Friedrich Maritsch,<sup>4</sup> Borislava Pavlova,<sup>4</sup> Valérie Philippon<sup>2</sup>

**Objective** Following guidance on data sharing from the International Committee of Medical Journal Editors,<sup>1</sup> Takeda requires the inclusion of a data sharing statement in most manuscripts reporting results from interventional clinical trials. The standard version of this statement indicates that deidentified patient data will be made available within 3 months to researchers who provide a methodologically sound proposal. The implementation of this policy for data sharing by request was assessed by reviewing published articles to

**Figure 24. Reason for Modified or Missing Data Sharing Statement**



determine whether a data sharing statement was included, modified, or not included, and the reason(s) for not sharing data were identified.

**Design** To identify articles for inclusion in this study, an internal publication management system was used. Articles published online between July 2020 and November 2021 reporting primary or secondary analyses from clinical trials sponsored by Takeda and reviewed by an internal clinical trial transparency (CTT) team were included in this analysis. Each article was assigned to 1 of the following categories: inclusion of standard data sharing statement, inclusion of modified data sharing statement, statement indicating there is no plan to share data, or no data sharing statement.

**Results** A total of 36 interventional clinical trial manuscripts reviewed by the CTT team were included in the analysis. Of these, 26 (72%) included a statement outlining data availability: 11 (31%) used the recommended standard data sharing statement as outlined in the internal policy, while 15 (42%) contained a modified statement on data availability, ranging from minor editorial amendments (generally consistent with the standard statement) to substantial differences in wording (due to specific journal requirements) (Figure 24). Only 2 articles (6%) contained a statement indicating that there was no plan to share the data—both were reporting results from rare disease studies involving very small numbers of patients, and the reason was concern regarding the risk of patient reidentification due to the number of patients. No data sharing statement was included in 8 articles (22%) for various reasons (Figure 24). Of the journals that published articles without a data sharing statement, 3 had policies requiring inclusion of a data sharing statement, while the others had policies allowing and/or encouraging inclusion of a data sharing statement.

**Conclusions** Sharing underlying data sets of published studies is important to promote transparency and facilitate new research. Following the adoption of a standardized approach to data sharing by request, most interventional clinical trial articles in this analysis contained a data sharing statement. In a few cases (n = 2), the statement indicated that

data would not be shared to protect patient privacy in situations in which deidentification of patient-level data was not possible.

## Reference

1. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Ann Intern Med.* 2017;167(1):63-65. doi:10.7326/M17-1028

<sup>1</sup>Takeda Pharmaceuticals International AG, Zürich, Switzerland; <sup>2</sup>Takeda Development Center Americas, Inc, Cambridge, MA, USA, jesse.potash@takeda.com; <sup>3</sup>Takeda Pharmaceuticals U.S.A., Inc, Lexington, MA, USA; <sup>4</sup>Baxalta Innovations GmbH, Vienna, Austria

**Conflict of Interest Disclosures** Colin McKinnon, Jesse Potash, Teodor G. Paunescu, Borislava Pavlova, and Valérie Philippon are current or past (V.P.) employees of Takeda and are Takeda stockholders. Callan Fromm is a consultant for the Global Publications Department at Takeda. Ingeborg Cil and Friedrich Maritsch are consultants for Takeda's Clinical Trial Transparency Department at Baxalta (a Takeda company). Hajin Yang is a postdoctoral fellow at Takeda.

**Funding/Support** This study was sponsored by Takeda Development Center Americas, Inc.

**Role of the Funder/Sponsor** The sponsor played a role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Diversity and Inclusion

### Enrollment and Representativeness in Contemporary Asthma Clinical Trials

Leslie L. Chang,<sup>1,2</sup> Clement D. Lee,<sup>1,3</sup> Katherine S. Takvorian<sup>1,2</sup>

**Objective** Asthma disproportionately affects historically marginalized racial and ethnic populations in the US.<sup>1</sup> However, these populations have been reported to be underrepresented in clinical trials,<sup>2</sup> limiting generalizability of study conclusions. The aim of this study was to examine enrollment and representativeness of study populations in contemporary asthma clinical trials.

**Design** A systematic search of the PubMed database was performed to identify randomized clinical trials enrolling 100 or more individuals aged 18 years or older with asthma living in the US that were published between January 1, 2015, and December 31, 2021. Exclusion criteria included trials in which asthma was not the major disease, international trials, and secondary analyses or meta-analyses of primary trials. Qualifying trials were abstracted for basic trial information, study design, and baseline epidemiological characteristics of the participants. To evaluate representativeness of study populations compared with Centers for Disease Control and Prevention epidemiological data for people with asthma, standardized mean differences using Cohen *d* effect size were calculated; underrepresentation and overrepresentation were defined as the proportions of enrolled participants that were 1

**Table 58. Baseline Characteristics of Randomized Clinical Asthma Trials, 2015 to 2021**

Characteristic	Trials, No. (%) (N = 35)
Participants, median (IQR), No.	311 (167-496)
<b>Sites</b>	
Single center	10 (28.6)
Multicenter	20 (57.1)
Unknown	5 (14.3)
<b>Intervention type</b>	
Drugs	17 (48.6)
Lifestyle related	18 (51.4)
Procedures	0
Mixed	0
<b>Source of support</b>	
Pharmaceutical industry	11 (31.4)
Government	14 (40.0)
Private or foundation	5 (14.3)
Academic	5 (14.3)
Undetermined	0
<b>Distribution reported</b>	
Age	33 (94.3)
Race and ethnicity	29 (82.9)
Sex	34 (97.1)
Educational level	12 (34.3)
Income level	10 (28.6)

or less SD and 1 or more SD of national proportions, respectively.

**Results** Of the 1032 trials screened, 35 met inclusion criteria (with the majority excluded due to international enrollment). The trials included a median of 311 participants (IQR, 167-496 participants). Approximately half of the trials were drug intervention trials, and the remainder investigated educational or behavioral interventions. There was a mix of funding sources, including pharmaceutical, government, academic, and private institutions. Almost all of the trials reported age and sex distribution of the participants. A total of 29 trials (82.9%) reported race or ethnicity. Only 12 (34.3%) and 10 (28.6%) trials reported educational and income levels, respectively (**Table 58**). With respect to representativeness of the study populations, among trials reporting participant sex, 1 of 33 (3.0%) had a proportion of females that was overrepresentative of the US population with asthma and no trials were underrepresentative. Among trials reporting Black race, 5 of 25 (20.0%) enrolled a proportion of Black participants that was overrepresentative of the US population with asthma and no trials were underrepresentative. Among trials reporting Hispanic ethnicity, 1 of 16 (6.25%) was overrepresentative and no trials were underrepresentative. Among trials reporting White race, 4 of 27 (14.8%) enrolled proportions that were

underrepresentative of the US population with asthma and no trials were overrepresentative.

**Conclusions** In contemporary clinical trials of adults with asthma in the US, the majority of trials reported participants' age, sex, and race and fewer reported education and income information. The study did not find that historically marginalized groups were systematically underrepresented in asthma clinical trials. Collecting and reporting patients' demographic and socioeconomic characteristics is crucial to understanding and generalizing trial results.

## References

1. Asthma and Allergy Foundation of America. Asthma disparities in America. Accessed February 25, 2022. <https://www.aafa.org/asthma-disparities-burden-on-minorities.aspx>
2. Downing NS, Shah ND, Neiman JH, Aminawung JA, Krumholz HM, Ross JS. Participation of the elderly, women, and minorities in pivotal trials supporting 2011-2013 US Food and Drug Administration approvals. *Trials*. 2016;17(1):199. doi:10.1186/s13063-016-1322-4

<sup>1</sup>New England Journal of Medicine, Boston, MA, USA, clee@nejm.org; <sup>2</sup>Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; <sup>3</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

## Women's Responses to Peer Review Invitations by 21 Biomedical Journals Prior to and During the COVID-19 Pandemic

Khaoula Ben Messaoud,<sup>1,2</sup> Sara Schroter,<sup>3</sup> Mark Richards,<sup>4</sup> Angèle Gayet-Ageron<sup>1,2</sup>

**Objective** Gender disparities have been shown to have increased during the COVID-19 pandemic at several levels in the research process, including in the submission and publication of biomedical research articles.<sup>1-3</sup> One study also reported an increase in the rate of women declining invitations to peer review biomedical research in the very early pandemic period.<sup>3</sup> However, women's response to peer review invitations during the longer term of the COVID-19 pandemic remains unknown. In this study, response to peer review invitations was explored using later COVID-19 data.

**Design** This was a retrospective cohort study of research manuscripts submitted to and sent for peer review at 21 biomedical journals from the BMJ Publishing Group from January 1, 2018, to May 31, 2021. Data were collected on the gender and geographic affiliation of the handling editor, the reviewers, and the last author; the journal's impact factor and peer review process (open vs anonymized); and the manuscript topic (COVID-19-related or not).<sup>1</sup> The primary outcome was response (agreed vs did not agree) to the review invitation. Multivariable mixed-effects logistic regression was performed with random factors on the intercept at journal and manuscript levels.

**Results** A total of 33,342 manuscripts were included; 257,025 reviewers were invited to review and 90,467 (35.2%) agreed. Mean (SD) number of invited and agreed reviewers by

manuscript were 7.7 (5.6) and 2.7 (1.2), respectively.

Multivariable analysis showed that agreement to review was associated significantly with time period and COVID-19 as a topic ( $P < .001$ ). Women agreed to review less frequently than men at any time period or COVID-19-related topic: odds ratios, 0.93 (95% CI, 0.90-0.95) in prepandemic, 0.82 (95% CI, 0.74-0.91) in January 2020 to January 2021 for COVID-19-related manuscripts, 0.89 (95% CI, 0.85-0.92) in January 2020 to January 2021 for non-COVID-19-related manuscripts, 0.81 (95% CI, 0.71-0.93) in February 2021 to February 2022 for COVID-19-related manuscripts, and 0.91 (95% CI, 0.86-0.97) in February 2021 to February 2022 for non-COVID-19-related manuscripts. Interaction between reviewers' gender, time period, or COVID-19 topic and agreement to review was significant ( $P = .049$ ). Women agreed significantly less often than men to review COVID-19-related manuscripts during the first year of the pandemic compared with the prepandemic period. Agreement to review was lower when the handling editor was a woman compared with a man: odds ratio, 0.96 (95% CI, 0.93-0.99) (**Table 59**).

**Conclusions** For both women and men reviewers, agreement to peer review was higher for COVID-19-related manuscripts compared with the prepandemic period. However, the representativeness of women researchers among invited reviewers was below parity. Moreover, women refused more frequently than men to participate in peer review. More generally, women's points of views need to be fostered to have diverse vision in research. Editors should ensure a fair parity in the invitation to peer review or increase the proportion of women invited to be able to obtain parity during peer review.

## References

1. Gayet-Ageron A, Ben Messaoud K, Richards M, Schroter S. Women authorship of COVID-19 research in manuscripts submitted to 11 biomedical journals: cross sectional study. *BMJ*. 2021;375:n2288.
2. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: global gender disparities in science. *Nature*. 2013;504(7479):211-3.
3. Squazzoni F, Bravo G, Grimaldo F, García-Costa D, Farjam M, Mehmani B. Gender gap in journal submissions and peer review during the first wave of the COVID-19 pandemic: a study on 2329 Elsevier journals. *PLOS ONE*. 2021;16(10):e0257919.

<sup>1</sup>Department of Health and Community Medicine, University of Geneva, Geneva, Switzerland, khaoula.benmessaoud@unige.ch;

<sup>2</sup>Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva, Switzerland; <sup>3</sup>The BMJ, London, UK; <sup>4</sup>Article Transfer Service, BMJ Publishing Group, London, UK

**Conflict of Interest Disclosures** The authors declare support from the University Hospitals of Geneva, Geneva Medical School, and the Swiss National Science Foundation for the submitted work. Sara Schroter and Mark Richards are employees of the BMJ Publishing Group. No other disclosures were reported.

**Funding/Support** This study was supported by a grant from the Swiss National Science Foundation (No. 192374).

**Table 59. Comparison of Reviewer, Journal, and Last Author Characteristics According to Agreement to Peer Review (n = 257,025) Manuscripts<sup>a</sup> Submitted to 21 BMJ Publishing Group Journals, January 1, 2018, to May 31, 2021: Multivariable Analyses<sup>b,c</sup>**

Characteristic	Odds ratio (95% CI)	P value	
Interaction between time period/topic and reviewer's gender		.049	
<b>Reviewer's gender in prepandemic</b>			
Men	1 [Reference]	<.001	
Women	0.93 (0.90-0.95)		
<b>Reviewer's gender by time period</b>			
<b>Jan 2020 to Jan 2021, COVID-19</b>			
Men	1 [Reference]	<.001	
Women	0.82 (0.74-0.91)		
<b>Jan 2020 to Jan 2021, no COVID-19</b>			
Men	1 [Reference]	<.001	
Women	0.89 (0.85-0.92)		
<b>Feb 2021 to Feb 2022, COVID-19</b>			
Men	1 [Reference]	<.001	
Women	0.81 (0.71-0.93)		
<b>Feb 2021 to Feb 2022, no COVID-19</b>			
Men	1 [Reference]	<.001	
Women	0.91 (0.86-0.97)		
<b>Last author gender</b>			
Men	1 [Reference]	.15	
Women	0.98 (0.96-1.01)		
<b>Journal impact factor</b>			
≤5	1 [Reference]	.001	
5-10	1.27 (0.96-1.67)		
>10	1.73 (1.29-2.32)		
<b>Type of peer review</b>			
Anonymized	1 [Reference]	<.001	
Open	0.43 (0.29-0.64)		
<b>Editor's gender</b>			
Men	1 [Reference]	.004	
Women	0.96 (0.93-0.99)		

<sup>a</sup>Manuscripts with 1 or more reviews completed.

<sup>b</sup>Model performed on completed values (n = 194,908).

<sup>c</sup>Mixed-effect logistic regression with manuscript and journal as random effects; model adjusted on reviewer, last author, and editor geographic institutional affiliation.

**Role of the Funder/Sponsor** The funder had no role in considering the study design or in the collection, analysis, or interpretation of data; the writing of the abstract; or the decision to submit the abstract for presentation.

### Assessment of Potential Barriers to Inclusion in Randomized Clinical Trials Published in Top General and Internal Medical Journals

Shelly Melissa Pranić,<sup>1,2</sup> Ksenija Baždarić,<sup>3</sup> Iván Pérez-Neri,<sup>4</sup> Maria Dulce da Mota Antunes de Oliveira Estevão,<sup>5</sup> Vinayak Mishra,<sup>6</sup> Joanne A. McGriff<sup>7</sup>

**Objective** Racial and ethnic minority groups are underrepresented in clinical research. Racially diverse individuals who speak languages other than English or have limited proficiency may be hindered from participation in randomized clinical trials (RCTs) through eligibility criteria.<sup>1,2</sup>

This study sought to assess English language requirements for enrollment in registered and published RCTs.

**Design** In a cross-sectional design, PubMed, Scopus, Epistemonikos, EBSCO Host, COVID-evidence, Web of Science Core Collection, and the World Health Organization COVID-19 databases were searched for RCTs in the top 10 first-quartile general and internal medicine journals in 2017 on May 4, 2022, with at least 1 US site comparing heart disease, stroke, cancer, asthma, influenza and pneumonia, diabetes, HIV/AIDS, and COVID-19 drug interventions with standard or usual care or placebo with ClinicalTrials.gov registration and protocols. Phrases collected from a previous assessment were searched for in the eligibility criteria in protocols and ClinicalTrials.gov records that indicated that English language was a requirement for trial enrollment. Good agreement was achieved by independent selection by 2 reviewers for inclusion ( $\kappa = 0.85$ ; 95% CI, 0.75-0.95) and data extraction and identification of language requirements in RCTs ( $\kappa = 0.98$ ; 95% CI, 0.87-1.00) from a sample of 50 RCTs. The primary outcome was the frequency of RCTs with English language requirements in eligibility criteria in protocols and ClinicalTrials.gov records by disease and funder type (industry funders had at least 1 industry funder, while nonindustry funders had no industry funding). Secondary outcomes were readability of eligibility criteria in ClinicalTrials.gov records and reporting of race as a demographic variable. Readability was assessed with Flesch-Kincaid grade (FKG) level (ranges from grades 0 to 18 [college graduate]) and Gunning-Fog (GF) (ranges from grades 0 to 20 [college graduate]), where lower grades correspond to easier readability. Mann-Whitney and Kruskal-Wallis tests compared readability between funder and disease with a 2-tailed P value set at less than .05.

**Results** A total of 39 of 2663 RCTs from *Annals of Internal Medicine* (n = 1), *JAMA* (n = 14), *JAMA Internal Medicine* (n = 3), *Lancet* (n = 12), *PLoS Medicine* (n = 1), and *New England Journal of Medicine* (n = 8) were found. The eligibility criteria made no explicit statements about English or any other language required for enrollment (**Table 60**) for American Indian participants (median [range], 7 [1-110]), Asian participants (median [range], 18 [1-836]), Black participants (median [range], 54 [4-2534]), Latinx participants (median [range], 83 [2-492]), and White participants (median [range], 264 [3-8715]). The median (IQR) FKG and GF levels by disease were 13.20 (11.80-13.90;  $P = .99$ ) and 13.80 (12.10-15.00;  $P = .66$ ), respectively. By funder, the median (IQR) FKG and GF levels were 13.20 (11.80-13.90;  $P = .16$ ) and 13.80 (12.10-15.00;  $P = .13$ ).

**Conclusions** Racial and ethnic minority groups were underrepresented in RCTs; there was low explicit reporting of required languages in RCT eligibility criteria; and readability levels of protocols were high. Trialists and researchers should be aware of the importance of the inclusion of underrepresented individuals, the explicit reporting of languages for participants, and the readability of trial information.

**Table 60. Reporting of English Language Requirements for Trial Enrollment in Protocol Eligibility Criteria of Randomized Clinical Trials**

Characteristic	No. (%)			
	Total (N = 39)	English language	Other language	No language requirement
<b>Disease studied</b>				
Heart disease	6 (15.4)	0	0	6 (15.4)
COVID-19	18 (46.1)	3 (7.7)	1 (2.6)	14 (35.3)
HIV/AIDS	5 (12.8)	0	1 (2.6)	5 (12.8)
Pneumonia	1 (2.6)	0	0	1 (2.6)
Diabetes	2 (5.1)	0	0	2 (5.1)
Cancer	5 (12.8)	1 (2.6)	1 (2.6)	3 (7.7)
Asthma	1 (2.6)	0	0	1 (2.6)
<b>Funder type</b>				
Industry	22 (56.4)	3 (13.6)	2 (9.0)	17 (77.3)
Nonindustry	17 (43.6)	1 (6.0)	1 (6.0)	15 (88.2)

## References

- George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *Am J Public Health.* 2014;104(2):e16-e31. doi:10.2105/AJPH.2013.301706
- Muthukumar AV, Morrell W, Bierer BE. Evaluating the frequency of English language requirements in clinical trial eligibility criteria: a systematic analysis using ClinicalTrials.gov. *PLoS Med.* 2021;18(9):e1003758. doi:10.1371/journal.pmed.1003758

<sup>1</sup>University of Split School of Medicine, Split, Croatia, shelly.pranic@mefst.hr; <sup>2</sup>Cochrane Croatia, Split, Croatia; <sup>3</sup>University of Rijeka Faculty of Medicine, Rijeka, Croatia; <sup>4</sup>National Institute of Neurology and Neurosurgery, Mexico City, Mexico; <sup>5</sup>Escola Superior de Saúde da Universidade do Algarve, Faro, Portugal; <sup>6</sup>University of Liverpool, Liverpool, UK; <sup>7</sup>Rollins School of Public Health, Emory University, Atlanta, GA, USA

**Conflict of Interest Disclosures** None reported.

## Editorial and Peer Review Process

### Results From a Preprint Review Opt-in Review Process at *eLife*

Emma Smith,<sup>1</sup> Andy Collings<sup>1</sup>

**Objective** In 2020, *eLife* piloted a review process in which peer reviews were posted publicly to preprints while manuscripts were simultaneously evaluated for publication. This pilot built on a 2018 *eLife* trial that published all reviewed submissions and found that a lower percentage of submissions were sent for peer review in the pilot process compared with the regular process.<sup>1,2</sup> An exploratory descriptive analysis was performed, comparing decision outcomes in the 2020 pilot with the regular process and

checking for outcome disparities between senior author demographic characteristics.

**Design** Between March 2020 and November 2020, authors could opt into the preprint review process during submission. In this process, authors committed to publicly posting the reviews alongside their preprint, irrespective of *eLife*'s decision. As an incentive, authors bypassed triage unless the editors found fundamental flaws. Data on submission outcomes were retrieved from *eLife*'s submission system. The senior author's demographic information was provided by the submitting author via an optional survey. The percentages of reviewed and accepted submissions were compared between the preprint review and the regular review processes. This comparison was also performed based on the senior author's gender and geographic region.

**Results** *eLife* received 420 preprint review and 6397 regular submissions. A higher percentage of preprint review submissions (345 [82.1%]) were reviewed relative to regular submissions (2057 [32.2%]). A higher percentage of preprint review submissions were also accepted (96 [22.9%]) compared with regular submissions (1071 [16.7%]). A slightly higher percentage of preprint review submissions from women senior authors were reviewed compared with men (**Table 61**). There was little difference in the percentage of submissions accepted based on the senior author's gender for regular submissions (**Table 61**). In both processes, the highest percentages of submissions reviewed and accepted were from senior authors in North America, and the lowest percentages were from Asia (**Table 61**). A higher percentage of submissions were accepted from Asia in the preprint review process relative to the regular process, although few senior authors from Asia submitted to preprint review (n = 23), so these results must be interpreted with caution.

**Conclusions** A higher percentage of preprint review submissions were reviewed and accepted, which suggests that improving access to peer review may provide a mechanism to mitigate bias in publishing. However, large differences in the number of submissions between the preprint review and regular processes and some small preprint review sample sizes limit the analysis that could be performed and conclusions that can be drawn. Future initiatives should take greater care to engage broader communities, especially in Asia. These results informed *eLife*'s shift to a "publish, then review" model of publishing.<sup>3</sup>

## References

- Eisen MB. Peer review: new initiatives to enhance the value of *eLife*'s process. Inside *eLife*. November 7, 2019. Accessed June 8, 2022. <https://elifesciences.org/inside-elife/e9091cea/peer-review-new-initiatives-to-enhance-the-value-of-elife-s-process>
- Smith E, Collings A. Peer review: first results from a trial at *eLife*. Inside *eLife*. January 7, 2019. Accessed June 8, 2022. <https://elifesciences.org/inside-elife/262c4c71/peer-review-first-results-from-a-trial-at-elife>

**Table 61. Number of Submissions Received, Reviewed, and Accepted in Preprint Review and Regular Review Processes**

Senior author demographic characteristic <sup>a</sup>	No. of preprint review submissions	No. of regular submissions	No. (%) of manuscripts reviewed		No. (%) of manuscripts accepted	
			Preprint review	Regular review	Preprint review	Regular review
<b>Gender</b>						
Men	153	3416	121 (79.1)	1180 (34.5)	36 (23.5)	627 (18.4)
Women	56	1345	49 (87.5)	442 (32.9)	13 (23.2)	233 (17.3)
<b>Region</b>						
Asia	23	927	15 (65.2)	191 (20.6)	4 (17.4)	76 (8.2)
Europe	88	1812	75 (85.2)	595 (32.8)	14 (15.9)	308 (17.0)
North America	83	1822	72 (86.8)	783 (43.0)	25 (30.1)	445 (24.4)

<sup>a</sup>Authors could self-declare their gender, but all preprint review authors declared either male or female gender identity, so results for other gender identities (eg, genderfluid, gender nonconforming, or nonbinary) are not shown in the table. Few preprint review submissions were received from Africa (n = 2), South America (n = 2), or Oceania (n = 10), so results from these regions are not shown. The demographic information survey was optional and only presented to authors who submitted via the eLife website, so the data for gender identity and region are incomplete.

3. Eisen MB, Akhmanova A, Behrens B, Harper DM, Weigel D, Zaidi M. Peer review: implementing a “publish, then review” model of publishing. *eLife*. 2020;9:e64910. doi:10.7554/eLife.64910

<sup>1</sup>eLife Sciences Publications, Cambridge, UK, e.smith@elifesciences.org

**Conflict of Interest Disclosures** None reported.

**Acknowledgment** The authors thank Peter Rodgers (*eLife*'s features editor) for comments and suggestions.

## A Survey of Authors' Experiences With Poor Peer Review Practices

Kyle McCloskey,<sup>1</sup> Jon F. Merz<sup>2</sup>

**Objective** A constructive peer review can bring a fresh perspective to help improve an author's work. Until recently, there has been little guidance on what constitutes a good or poor peer review and no empirical study about researchers' experiences with unhelpful or upsetting reviews.<sup>1-3</sup> This pilot study aimed to explore the types of poor peer review practices (PPRP) experienced by authors and assess their association with authors' ability to disseminate research.

**Design** An anonymous 51-question survey was designed through Qualtrics. An invitation to complete the survey was emailed to a random sample of 500 researchers funded by the National Institutes of Health in 2018 (with the replacement of undeliverable email addresses) and posted to a bioethics discussion forum (mcw-bioethics@mailman.mcw.edu) with approximately 600 members. Three mailings were performed between April 5 and April 19, 2022. The study was determined to be exempt by the University of Pennsylvania institutional review board. A 28-item list of PPRP was developed following a literature review; respondents' experiences were assessed with yes-no and open-ended questions. The PPRP was assessed through author responses to Likert scale questions. Exploratory nonparametric analyses of respondents' reported experiences and demographic characteristics were performed using Stata, version 12.1 (StataCorp LLC). Data were reported according to the Consensus-Based Checklist for Reporting of Survey Studies (CROSS).

**Results** A total of 112 researchers completed the surveys, approximately 10% of those solicited. Respondents were predominantly male (59 of 109 [54%]), held a PhD (87 of 112 [78%]), were older than 50 years (73 of 111 [66%]), identified as White race (92 of 106 [87%]), published more than 50 peer-reviewed papers in their career (71 of 111 [64%]), were trained in the humanities or social sciences (61 of 112 [54%]), and conducted primarily empirical research (68 of 111 [61%]). The mean number of PPRP experienced per author was 12.5 of 28 (44.6%) (range, 0-27; 95% CI, 11.2-13.8). A total of 57% (63 of 111) of authors admitted abandoning a manuscript after receiving unfair peer reviews; 67% (74 of 111) of authors sometimes or often received insightful peer reviews that improved the quality of their final papers. An exploratory univariate analysis of the association between the total number of PPRP reported per author and demographic characteristics and reported positive peer review experiences is presented in **Table 62**. This analysis suggests areas for future study, particularly the need to explore types of helpful reviews.

**Conclusions** This pilot study was the first to date to assess researchers' experiences with PPRP. The low response rate suggests high selectivity and bias and that respondents report experiences with an extensive range of PPRP. The study lays the groundwork for future research, which may be valuable for improving peer review quality.

## References

- Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med*. 2006;99(4):178-182. doi:10.1258/jrsm.99.4.178
- Mavrogenis AF, Quaile A, Scarlat MM. The good, the bad and the rude peer-review. *Int Orthop*. 2020;44(3):413-415. doi:10.1007/s00264-020-04504-1
- Silbiger NJ, Stubler AD. Unprofessional peer reviews disproportionately harm underrepresented groups in STEM. *PeerJ*. 2019;7:e8247. doi:10.7717/peerj.8247

<sup>1</sup>Drexel University College of Medicine, Philadelphia, PA, USA;

<sup>2</sup>Perelman School of Medicine, University of Pennsylvania, Department of Medical Ethics and Health Policy, Philadelphia, PA, USA, merz@pennmedicine.upenn.edu

**Table 62. Univariate Exploratory Analysis of Total PPRP**

Factor (No. of respondents)	PPRP, mean (SD), No. per author	z Score	P value
<b>Age range, y</b>			
20-29 (9)	15.4 (5.9)	-2.51	.01 <sup>a</sup>
30-39 (29)	14.7 (7.0)		
40-49 (33)	12.0 (7.0)		
50-59 (40)	10.8 (6.1)		
<b>Gender</b>			
Female (50)	11.5 (6.9)	1.39	.16 <sup>b</sup>
Male (59)	13.3 (6.6)		
<b>Race</b>			
White (92)	12.0 (7.0)	1.62	.10 <sup>b</sup>
Non-White (14)	15.0 (4.9)		
<b>Underrepresented group</b>			
No (84)	11.8 (7.1)	1.97	.05 <sup>b</sup>
Yes (22)	14.8 (5.2)		
<b>Career peer-reviewed publications, No.</b>			
≤10	10.8 (5.8)	2.37	.02 <sup>a</sup>
11-50	11.1 (6.8)		
51-100	11.9 (6.9)		
>100	14.7 (6.6)		
<b>Have received positive helpful peer reviews</b>			
Never (3)	16.0 (5.0)	-2.53	.01 <sup>a</sup>
Rarely (34)	14.0 (7.3)		
Sometimes (65)	12.3 (6.4)		
Often (9)	7.7 (5.0)		

Abbreviation: PPRP, poor peer review practices.

<sup>a</sup>Cuzick extension of Wilcoxon rank sum test for ordered groups.<sup>b</sup>Nonparametric Wilcoxon rank sum (Mann-Whitney) test.**Conflict of Interest Disclosures** None reported.

### Using Custom Questions to Assess Patient Involvement in Articles Submitted to a General Medical Journal

Victoria Saigle,<sup>1</sup> Meredith Weinhold,<sup>1</sup> Kirsten Patrick,<sup>1</sup> Andreas Laupacis<sup>1,2</sup>

**Objective** The *Canadian Medical Association Journal* (*CMAJ*) has committed to increasing patient engagement in its content (“patient” includes patients, families, and caregivers).<sup>1</sup> As part of its evaluation strategy, *CMAJ* added custom questions to its submission platform to monitor submissions with patient engagement. This study aimed to assess how authors responded to these custom questions and the outcomes of submissions with patient engagement.

**Design** This cross-sectional study analyzed *CMAJ* submissions between March 2, 2021, and March 1, 2022, in which authors responded affirmatively to either of the following questions automatically presented on submission: (1) “Does your submission have patient (co-)authors or

acknowledge the contributions of patient, family, or caregiver partners? Select ‘no’ if patients were only research participants or described as clinical cases” (mandatory for all articles) and (2) “Was any part of your research study design or conduct informed by patients directly (eg, through patient involvement on your team)? If yes, please include these details in your manuscript and complete the GRIPP2 checklist (short form)” (asked of research articles only). Submissions were deemed to have patient involvement if a reviewer confirmed that 1 of the authors was a patient or that patient contributions were noted in the cover letter, open text fields, GRIPP2 short-form questionnaire, editor notes, or manuscript files. Submissions to the Research, Guideline, Clinical Review, Commentary, Humanities, Analysis, and Practice sections were included. The final sample was compared with a list of articles with patient involvement published by *CMAJ* during this time frame.

**Results** Among 1879 submissions, 173 of 1879 submitting authors (9.2%) indicated involvement of patient authors or partners, and 19 of 1879 (1.0%) indicated that patients had informed the work but were not authors or partners. Among these 192 submissions, 67 (34.9%) satisfied the journal’s criteria for patient engagement (**Table 63**). This included 33 of 80 research submissions (41.3%). Sixty-three of 67 submissions (94.0%) deemed to have patient engagement were submitted by authors with Canadian affiliations. Overall, 19 of 63 submissions (30.2%) that responded affirmatively to 1 of these questions and were assessed to have patient engagement were published; among these, 11 of 19 (57.9%) had a patient author. In comparing this list with known publications with patient involvement, an additional 14 articles were identified that did not respond affirmatively to the 2 questions under study.

**Table 63. CMAJ Submissions Indicating Patient Involvement March 2, 2021, to March 1, 2022**

Variable	Total	Only Q1 <sup>a</sup>	Only Q2 <sup>b</sup>	Q1 <sup>a</sup> and Q2 <sup>b</sup>
Author answered affirmatively	192	141	19	32
Met <i>CMAJ</i> criteria for patient involvement	67	35	5	27
Corresponding author affiliation Canada	63	33	5	25
Accepted	19	17	0	2
Commentary	1	1	0	0
Guidelines	1	1	0	0
Humanities	11	11	0	0
Practice	4	4	0	0
Research	2	0	0	2
At least 1 author was a patient	11	9	0	2

Abbreviation: *CMAJ*, Canadian Medical Association Journal.<sup>a</sup>Q1 represents submission question “Does your submission have patient (co-)authors or acknowledge the contributions of patient, family, or caregiver partners? Select no if patients were only research participants or described as clinical cases.”<sup>b</sup>Q2 represents submission question “Was any part of your research study design or conduct informed by patients directly (eg, through patient involvement on your team)? If yes, please include these details in your manuscript and complete the GRIPP2 checklist (short form).

**Conclusions** Most authors incorrectly answered the questions pertaining to patient involvement. Although custom questions may be useful in helping journal staff identify patient engagement in submitted manuscripts, their utility may depend on the submitting authors' familiarity with terms related to patient engagement. Future work should seek to improve the clarity of these questions.

## Reference

1. Saigle V, Miller J, Dumez V, Patrick K. Embedding patient voices in *CMAJ*. *CMAJ*. 2021;193(27):E1046–E1047.

<sup>1</sup>CMAJ, Ottawa Hospital Research Institute, Ottawa, ON, Canada, victoria.saigle@cmaj.ca; <sup>2</sup>Dalla Lana School of Public Health and Faculty of Medicine, University of Toronto, Toronto, ON, Canada

**Conflict of Interest Disclosures** The authors are all employed by the *Canadian Medical Association Journal*.

## Education/Training

### Developing the Next Generation of Editors and Reviewers Through a Trainee-Led Editorial Board in Neurology

Roy E. Strowd,<sup>1</sup> Whitley W. Aamodt,<sup>2</sup> Ariel M. Lyons-Warren,<sup>3</sup> Kathleen M. Pieper,<sup>4</sup> José G. Merino<sup>5</sup>

**Objective** Since 2004, the Resident & Fellow Section (RFS) of *Neurology* has provided an outlet for neurology trainees and educators to publish manuscripts. The RFS board members are trainees selected annually for a 3-year term to serve as reviewers and learn publication science. The objective of this study was to assess the association of RFS board membership with (1) future leadership roles at academic journals, (2) peer review statistics, and (3) authorship.

**Design** This retrospective cohort study was conducted with an embedded, nested case-control study. A cohort of all former RFS editorial board members was identified. For the nested component, cases consisted of editorial board members from 2018 to 2020, whereas controls were unselected applicants who made the penultimate round of candidacy. Demographic, reviewer, and authorship data including reviewer invitation acceptance rate, review completion rate (total number and past 12 months), review turnaround time, number of manuscripts submitted to *Neurology*, and acceptance rate in the 3 years following board application were collected via review of the BenchPress online submission system. Gender identity and whether each board member held a future leadership position in *Neurology* or other academic journals was collected via survey. Descriptive statistics were performed; unpaired *t* tests were used to compare peer review statistics between cases and controls.

**Results** From 2004 to 2021, 77 trainees served on the editorial board; 50 were male (65%) and 7 were international (9%). Of this cohort, 31 (40%) served in 65 future editorial leadership roles at academic journals, including 20 who were editorial board members (**Table 64**), 14 who were RFS

mentor-mentee peer review program mentors, 19 section editors, 2 contributing/guest editors, 7 associate editors, and 3 editors in chief. In addition, 41 (54%) continued to review for the RFS in 2021 and completed 291 reviews (mean [SD], 3.8 [7.0] reviews per graduate). In the past 3 years, 167 residents applied to the editorial board (mean [SD], 56 [10] applications per year). Compared with 23 controls who were not selected, 20 cases completed significantly more peer reviews over the following 3 years (mean [SD], 51.3 [21.0] vs 7.7 [12.0]; *P* < .001). Mean (SD) review turnaround time (cases, 7.2 [2.9] days vs controls, 6.7 [4.0]; *P* = .70), number of manuscripts authored (7.3 [9.0] vs 3.7 [4.0]; *P* = .11), number of manuscripts accepted (2.9 [3.0] vs 1.3 [2.0]; *P* = .09), and acceptance rate (43% vs 30%; *P* = .21) did not differ significantly.

**Conclusions** A trainee-led editorial board prepares medical trainees to review manuscripts, publish content, and develop future passion in journalology. Although the association of board participation with authorship was not different from highly qualified controls and expectations of board membership is an important confounder, engagement in more peer review may promote interest in publication science.

<sup>1</sup>*Neurology*, Wake Forest School of Medicine, Charlotte, NC, USA, rstrwod@wakehealth.edu; <sup>2</sup>*Neurology*, University of Pennsylvania, Pittsburgh, PA, USA; <sup>3</sup>Department of Pediatrics, Section of Pediatric Neurology and Developmental Neuroscience, Baylor College of Medicine, Houston, TX, USA; <sup>4</sup>*Neurology*; <sup>5</sup>Georgetown University Medical Center, Washington, DC, USA

**Conflict of Interest Disclosures** All authors are affiliated with the *Neurology* Resident & Fellow Section. Whitley W. Aamodt receives an editorial stipend as the deputy section editor. Ariel M. Lyons-Warren serves as chair of the Resident & Fellow Section Mentored Review Program, and Roy E. Strowd serves as the section editor. Kathleen M. Pieper is the senior managing editor of *Neurology*, and José G. Merino serves as editor in chief of the *Neurology* journals. Roy E. Strowd receives an editorial stipend as section editor and serves as a consultant for Monteris Medical Inc and Novocure; he has received stipends as an educational lecturer for Lecturio and Kaplan and has received research/grant support from the American Academy of Neurology, the American Board

**Table 64. Future Editorial Leadership Roles for Graduates of the Neurology Resident & Fellow Section (RFS) Editorial Board**

Type of editorial leadership role	Respondents who have held this role, No. (%)	Examples of leadership role titles
Editorial board member	20 (74)	<i>Neurology: Clinical Practice, Stroke, Neuro-hospitalist, Epileptic Disorders</i>
Section editor	19 (70)	<i>Neurology Disputes &amp; Debates, Journal of Neuro-ophthalmology</i>
RFS peer review mentor	14 (52)	
Associate editor	7 (26)	<i>Brain &amp; Life, BMC Research Notes, Arquivos de Neuro-Psiquiatria</i>
Editor in chief	3 (11)	<i>Neurology: Clinical Practice, Headache</i>
Contributing editor	2 (7)	Guest editor for JOVE
None	16 (59)	

of Psychiatry and Neurology, the American Society for Clinical Oncology, the Southeastern Brain Tumor Foundation, and Jazz Pharmaceuticals. Kathleen M. Pieper and José G. Merino also receive compensation for their roles. No other disclosures were reported.

**Acknowledgments** We thank our editorial team and peer reviewers for their time and commitment to the *Neurology* Resident & Fellow Section.

## Ethics and Ethical Concerns

### A Computational Method to Address Strategic Behavior in Peer Review

Komal Dhull,<sup>1</sup> Steven Jecmen,<sup>1</sup> Pravesh Kothari,<sup>1</sup> Nihar B. Shah<sup>1</sup>

**Objective** In many peer review processes, such as grant proposal review and conference paper review in computer science, only a fixed number of submissions are accepted. Moreover, many authors are also tasked with reviewing other submissions. This is known to lead to strategic behavior, whereby reviewers manipulate the reviews they provide to increase the chances of their own submissions getting accepted.<sup>1,2</sup> The objective of this work was to prevent such unethical behavior in peer review.

**Design** New computational methods to address this problem were developed. The methods build on a prior method for strategy-proof reviewer assignment by Alon et al.<sup>3</sup> Their method randomly partitions reviewers into 2 groups and assigns reviewers to review papers authored by reviewers in the other group. Their method then accepts for publication an equal number of papers from each group, thus guaranteeing that no reviewer can influence the outcome of their own papers by manipulating the reviews they provide (ie, “strategyproofness”). The methods proposed in the present work more carefully choose the partition of reviewers to maximize an assignment quality objective, while still satisfying strategyproofness. Large venues frequently consider such an assignment quality objective when using artificial intelligence to assign reviewers.<sup>2</sup> The assignment procedure first computes a similarity score between every reviewer-paper pair as a proxy for assignment quality and then assigns reviewers to papers in a manner that maximizes the cumulative similarity score of the assigned reviewer-paper pairs subject to load constraints. The proposed methods aim to choose the highest-similarity assignment subject to the strategyproofness guarantee. The strategyproofness constraint could reduce the cumulative similarity score of the assignment, and this metric was empirically evaluated. This evaluation was performed on data from the International Conference on Representation Learning 2018, a top conference in artificial intelligence that reviewed 911 full papers and was a terminal venue of publication. The optimal cumulative similarity score of the assignment in the absence of strategyproofness was computed and compared with that obtained under the proposed algorithms, as well as the aforementioned baseline algorithm.<sup>3</sup>

**Results** Figure 25 displays the reduction in the cumulative similarity of the assignment produced by the proposed algorithms. For reviewer and paper loads of 1, 2, and 3, respectively, the cycle-breaking algorithm lost only 3.27%, 4.73%, and 5.80% of optimal similarity; the coloring algorithm lost 11.1%, 11.9%, and 11.7% of optimal similarity; and the baseline random algorithm lost 19.0%, 17.1%, and 15.9% of optimal similarity. The similarity loss of the cycle-breaking algorithm was at least 2.5 times less than that of the random algorithm.

**Conclusions** The proposed methods realized strategyproofness without a large reduction in cumulative similarity, indicating that assignment quality remained high.

### References

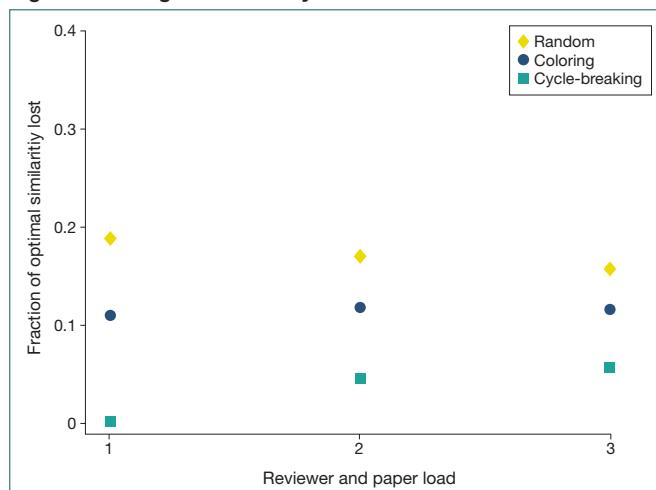
1. Bialiotti S, Robert G, Dirk H. Peer review and competition in the Art Exhibition Game. *Proc Nat Acad Sci*. 2016;113(30):8414-8419. doi:10.1073/pnas.1603723113
2. Shah N. An overview of challenges, experiments, and computational solutions in peer review. Communications of the Association for Computing Machinery. Posted online July 7, 2022. <https://www.cs.cmu.edu/~nihars/preprints/SurveyPeerReview.pdf>
3. Alon N, Fischer F, Procaccia A, Tennenholtz M. Sum of us: strategyproof selection from the selectors. In: *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, 2011. doi.org/10.1145/2000378.2000390

<sup>1</sup>Carnegie Mellon University, Pittsburgh, PA, USA, nihars@cs.cmu.edu

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported in part by the US National Science Foundation Career award 1942124 and a Google Research Scholar Award, both of which support research on the fundamentals of learning from people with applications to peer review.

**Figure 25. Assignment Quality Evaluation**



Cumulative similarity scores (proxy for assignment quality) of the proposed strategyproof algorithms (cycle-breaking and coloring) and a baseline strategyproof algorithm (random by Alon et al<sup>3</sup>) compared with the optimal cumulative similarity in the absence of strategyproofness. The y-axis represents the reduction in the cumulative similarity due to strategyproofness.

**Role of the Funder/Sponsor** The sponsors did not play a direct role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; or decision to submit the abstract for presentation.

**Additional Information** The code is available at [https://github.com/sjecmen/optimal\\_strategyproof\\_assignment](https://github.com/sjecmen/optimal_strategyproof_assignment).

## Funding/Grant Peer Review

### Assessment of Performance of Grant Peer Reviewers in the Canadian Health Research Funding System, 2019-2021

Clare L. Ardern,<sup>1,2</sup> Nadia Martino,<sup>3</sup> Sammy Nag,<sup>3</sup> Adrian Mota,<sup>4</sup> Karim M. Khan<sup>1,5,6</sup>

**Objective** Funders must make evidence-informed decisions about how best to continue to deliver fair, equitable, and inclusive (ie, high-quality) grant peer review. The Canadian Institutes of Health Research (CIHR) commenced a quality assurance program in fall 2019 to routinely monitor the quality of peer reviewer participation in and contributions to its Project Grant Competition Peer Review Panels. The aim was to describe the performance of grant peer reviewers in the Canadian health research funding system based on assessments made by Peer Review Panel leaders.

**Design** All Peer Review Panel chairs and scientific officers who led peer review for the CIHR Project Grant Competition rounds in 2019 (fall), 2020 (fall), and 2021 (spring and fall) completed CIHR's reviewer quality feedback form immediately after the Peer Review Panel meeting and returned the form to CIHR College of Reviewers staff. The form addressed 15 elements to characterize quality peer review across 5 domains (**Table 65**). All Peer Review Panel members consented to anonymous assessment of their performance by the chair or scientific officer of the panel. Chairs and scientific officers, supported by CIHR staff, used the feedback form<sup>1</sup> (5-7 minutes to complete per panel) to (1) assess reviews for appropriateness, robustness, and utility; (2) judge whether reviewer participation was professional, responsive, and engaged; and (3) identify peer reviewers with potential to serve as future chairs, scientific officers, or peer reviewer mentors.

**Results** The performance of 4438 peer reviewers (1828 female reviewers [41%], 2601 male reviewers [59%], and 9 reviewers [0.2%] who did not declare sex; 2459 unique peer reviewers participating across 4 competition rounds) was evaluated by 478 chairs and scientific officers. There were between 57 and 61 panels and approximately 1000 peer reviewers in each competition round. There were 1190 reviewers participating in 1 Project Grant Competition round, 698 reviewers participating in 2 rounds, 434 reviewers participating in 3 rounds, and 139 reviewers participating in all 4 rounds. Peer review for Project Grant Competitions in 2020 and 2021 was delivered online via Microsoft Teams. Approximately 1 in 3 peer reviewers was considered to have submitted outstanding reviews or participated constructively

**Table 65. Summary of Peer Review Quality Indicators**

Indicator	Reviewers, No. (%) <sup>a</sup>			
	Fall 2019 (n = 991) <sup>b</sup>	Fall 2020 (n = 1123) <sup>c</sup>	Spring 2021 (n = 1230) <sup>d</sup>	Fall 2021 (n = 1094) <sup>e</sup>
<b>Performance</b>				
Undertook additional tasks	51 (5.1)	23 (2.0)	47 (3.8)	39 (3.6)
Discussed additional applications	309 (31.2)	430 (38.3)	380 (30.9)	311 (28.4)
Outstanding review	285 (28.8)	487 (43.4)	386 (31.4)	356 (32.5)
<b>Potential</b>				
Panel chair	136 (13.7)	172 (15.3)	141 (11.5)	112 (10.2)
Panel scientific officer	171 (17.3)	196 (17.5)	171 (13.9)	171 (15.6)
Peer reviewer mentor	145 (14.6)	209 (18.6)	236 (19.2)	160 (14.6)
<b>Review quality</b>				
Lacks robustness	53 (5.3)	59 (5.3)	59 (4.8)	46 (4.2)
Lacks appropriateness	1 (0.1)	6 (0.5)	19 (1.5)	16 (1.5)
<b>Participation</b>				
Low participation	23 (2.3)	23 (2.0)	19 (1.5)	9 (0.8)
Major presentation weakness	27 (2.7)	20 (1.8)	14 (1.1)	8 (0.7)
Difficult to chair	9 (0.9)	24 (2.1)	11 (0.9)	9 (0.8)
Lacks professionalism	14 (1.4)	8 (0.7)	9 (0.7)	6 (0.5)
<b>Responsiveness</b>				
Late submitting review	24 (2.4)	22 (2.0)	3 (0.2)	2 (0.2)
Follow-up to submit scores or review	28 (2.8)	67 (6.0)	49 (4.0)	56 (5.1)
Follow-up for conflict of interest declaration	2 (0.2)	9 (0.8)	9 (0.7)	16 (1.5)

<sup>a</sup>N values are the number of peer reviewers who contributed to the competition peer review.

<sup>b</sup>Missing data for 56 peer reviewers.

<sup>c</sup>Missing data for 23 peer reviewers.

<sup>d</sup>Missing data for 51 peer reviewers.

<sup>e</sup>Missing data for 34 peer reviewers.

in discussions of additional applications not assigned to that reviewer (**Table 65**). At least 1 in 10 peer reviewers demonstrated potential as a future chair, scientific officer, or peer reviewer mentor (**Table 65**). At most, 1 in 20 peer reviewers was considered to have not performed adequately with respect to review quality, participation, or responsiveness.

**Conclusions** The quality of peer review for Project Grant Competitions at Canada's health research funding agency (as assessed with CIHR's Reviewer Quality Feedback Form) was consistent during the study period. Among peer reviewers, approximately 95% met the standard expected by chairs and scientific officers for review quality, participation, and responsiveness.

### Reference

- Canadian Institutes of Health Research. Reviewer quality feedback form. Accessed June 24, 2022. <https://cihr-irsc.gc.ca/e/51654.html>

<sup>1</sup>Department of Family Practice, University of British Columbia, Vancouver, BC, Canada, clare.ardern@ubc.ca; <sup>2</sup>Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Australia; <sup>3</sup>Canadian Institutes of Health Research College of Reviewers, Ottawa, ON, Canada; <sup>4</sup>Canadian Institutes of Health Research, Research Programs-Operations, Ottawa, ON, Canada; <sup>5</sup>School of Kinesiology, University of British Columbia, Vancouver, BC, Canada; <sup>6</sup>Canadian Institutes of Health Research-Institute of Musculoskeletal Health and Arthritis, Vancouver, BC, Canada

**Funding/Support** This work was supported by a Canadian Institutes of Health Research (CIHR) Research Operating Grant (Scientific Directors) held by Karim M. Khan.

**Role of the Funder/Sponsor** The funder (CIHR) had no role in the design and conduct of the study. CIHR's College of Reviewers coordinated data collection, management, and analysis as part of its mandate to foster and deliver high-quality peer review for health research in Canada. CIHR did not participate in interpretation of the data or the preparation of the abstract; CIHR reviewed and approved the abstract prior to submission. CIHR did not participate in the decision to submit the abstract for presentation.

**Conflict of Interest Disclosures** Nadia Martino (executive director, CIHR College of Reviewers), Sammy Nag (senior advisor, CIHR College of Reviewers), and Adrian Mota (associate vice president, CIHR Research Programs—Operations) are employees of CIHR. Nadia Martino and Sammy Nag are responsible for coordinating the Research Quality Assurance program for CIHR. Karim M. Khan is the CIHR—Institute of Musculoskeletal Health and Arthritis scientific director.

**Additional Contributions** Zakhirullah Sanaye (advisor, CIHR College of Reviewers), Nicole Haddadian (lead, CIHR College of Reviewers), and Simon Partridge (senior advisor, CIHR College of Reviewers) constructed the project database and helped produce the summary data.

**Additional Information** Karim M. Khan is a co-corresponding author.

## Instructions for Authors

### Analysis of Biomedical Journals' Instructions to Authors and Reviewers on Use of Reporting Guidelines

Peiling Wang,<sup>1</sup> Dietmar Wolfram<sup>2</sup>

**Objective** Reporting guidelines for biomedical publications have been developed for more than 2 decades. The extent of their appropriate use by authors has varied across journals.<sup>1,2</sup> Researchers found varied instructions for authors following the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline across journals.<sup>3</sup> To improve the quality of biomedical publications, it is important to bridge the gaps in adopting and implementing reporting guidelines. This study investigated which biomedical journals published the current major reporting guidelines and how they instructed authors and reviewers to adhere to the reporting guidelines.

**Design** Data were collected between May 2021 and June 2021 from the Enhancing the Quality and Transparency of Health Research (EQUATOR) network, the Reporting Guidelines websites (eg, <http://www.consort-statement.org/> and <https://prisma-statement.org/>), journals that published

the guidelines, and journals' instructions for authors and reviewers. From the 11 EQUATOR network–listed guidelines for main study types, there were 8 available guideline websites and 55 journals that published 1 or more current reporting guidelines. Additional data were collected from the 55 journals regarding endorsement of the specific guidelines and the journal's instructions for authors and reviewers, respectively. The data excluded instructions that only mentioned EQUATOR.

**Results** The journals that endorsed specific guidelines ranged from 24 for Standards for Quality Improvement Reporting Excellence (SQUIRE) reporting guideline–friendly journals to 1047 for Animal Research: Reporting of In Vivo Experiments (ARRIVE). However, only 55 journals published 1 to 10 guidelines (mean [SD], 6.3 [3.0]) (Table 66). Some journals published specific guidelines but did not give instructions regarding these guidelines; other journals did not publish specific guidelines but included them in instructions for authors and reviewers. Of the 55 journals, 40 (72.7%) provided instructions for authors, mentioning between 1 and 8 guidelines. Only 9 journals (16.4%) instructed reviewers about 5 guidelines. The guidelines in instructions for authors occurred more in open access journals (median, 5) than non–open access journals (median, 2). The collaborations among guidelines showed that the same contributors were involved in the development of multiple guidelines. For example, the Strengthening the Reporting of Observational

**Table 66. Journals That Published Reporting Guidelines and Their Instructions for Authors and Reviewers<sup>a</sup>**

Reporting guidelines	Website, Y/N	Journals that published the guidelines, No. (n = 55)	Journal instructions (n = 40) <sup>b</sup>		Proportion of instructions for reviewers vs authors, %
			For authors	For reviewers	
AGREE II	Y	3	2	0	0
ARRIVE	Y	7	20 <sup>c</sup>	1	5.0
CARE	Y	7	8 <sup>c</sup>	0	0
CHEERS	N	10	11 <sup>c</sup>	0	0
CONSORT	Y	9	29 <sup>c</sup>	8	27.6
PRISMA	Y	5	22 <sup>c</sup>	4	18.2
SPIRIT	Y	6	6	0	0
SQUIRE	Y	10	8	0	0
SRQR <sup>c</sup>	N	1	0	0	0
STARD	N	3	14 <sup>c</sup>	1	7.1
STROBE	Y	8	19 <sup>c</sup>	3	15.8
Mean (SD)	NA	6.3 (3.0)	13.9 (9.0) <sup>c</sup>	1.7 (2.5)	6.7 (9.6)

Abbreviations: AGREE II, Appraisal of Guidelines for Research & Evaluation Instrument; ARRIVE, Animal Research: Reporting of In Vivo Experiments; CARE, Case Report; CHEERS, Consolidated Health Economic Evaluation Reporting Standards; CONSORT, Consolidated Standards of Reporting Trials; EQUATOR, Enhancing the Quality and Transparency of Health Research; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; SQUIRE, Standards for Quality Improvement Reporting Excellence; SRQR, Standards for Reporting Qualitative Research; STARD, Standards for Reporting of Diagnostic Accuracy; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

<sup>a</sup>The EQUATOR network listed 11 reporting guidelines for main study types.

<sup>b</sup>Includes journals that mentioned reporting guidelines but did not publish them.

<sup>c</sup>SRQR was published in 1 journal, which did not mention any guidelines for authors or reviewers.

Studies in Epidemiology (STROBE) reporting guideline website referenced the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline, SQUIRE, and STARD. The level of implementation of adopted guidelines also varied by journal; answering yes or no questions about a relevant guideline, requiring authors to attach a guideline checklist, or citing guidelines.

**Conclusions** There is varied adoption and implementation of published guidelines by medical journals. Open access journals that published guidelines were more likely to instruct authors about the guidelines. The varied levels of instructions for authors and almost no instructions for reviewers on adherence to reporting guidelines call for further study. It should be of concern if a journal publishes reporting guidelines but does not instruct authors to adhere to the guidelines. Both authors and reviewers can contribute to improving the quality of biomedical research publications through adherence to established reporting guidelines.

## References

1. Caulley L, Catalá-López F, Whelan J, et al. Reporting guidelines of health research studies are frequently used inappropriately. *J Clin Epidemiol.* 2020;122:87-94. doi:10.1016/j.jclinepi.2020.03.006
2. Caulley L, Cheng W, Catalá-López F, et al. Citation impact was highly variable for reporting guidelines of health research: a citation analysis. *J Clin Epidemiol.* 2020;127:96-104. doi:10.1016/j.jclinepi.2020.07.013
3. Smidt N, Overbeke J, de Vet H, Bossuyt P. Endorsement of the STARD Statement by biomedical journals: survey of instructions for authors. *Clin Chem.* 2007;53(11):1983-1985. doi:10.1373/clinchem.2007.090167

<sup>1</sup>School of Information Sciences, University of Tennessee-Knoxville, Knoxville, TN, USA, peilingw@utk.edu; <sup>2</sup>School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

**Conflict of Interest Disclosures** None reported.

**Additional Information** Dietmar Wolfram is a co-corresponding author.

## Misconduct

### Experience With Communications to Medical Journals Requesting Investigation Into Published Articles With Possible Data Fabrication

Ben W. Mol,<sup>1,2</sup> Jim Thornton,<sup>3</sup> Wentao Li<sup>1</sup>

**Objective** When data integrity concerns are raised about a published article, the Committee of Publications Ethics (COPE) recommends that editors investigate. It is unknown how effective this process is.

**Design** A prospective cohort study on published articles with concerns about data fabrication was performed. Concerns could be multiple per study and included implausible timelines, implausible effect sizes, discrepancies between

publication and trial registration, copying of tables from other articles, plagiarism, authors having published fabricated studies elsewhere, and wrong statistics. Editors of the involved journals were contacted by email with a summary of the concerns for each article and asked the editor to investigate according to COPE recommendations. The email suggested that the editors request original data and included an offer to help investigate data sets if needed. The editors' responses, final decisions, and reactions are reported herein.

**Results** Between March 2017 and May 2022, editors were contacted about 546 articles (356 randomized clinical trials [65%]) by 105 authors from 6 countries (11 in 2017-2019; 56 in 2020; 222 in 2021; and 257 in 2022). Articles were mainly from the field of obstetrics/gynecology but also from urology, pediatrics, and infectious disease (COVID-19) and were published in 74 different journals (1 to 55 concerns per journal) by 5 different publishers. A total of 271 concerns (79%) were answered with confirmation of receipt. Most editors respected confidentiality, but some editors copied the complainant while writing to authors and 1 published an Editorial naming the complainant. Some suggested a Letter to the Editor. Five editors (7%) shared original data provided by the authors with the complainant for a total of 18 data sets, 14 (78%) of which showed signs of possible fabrication, mostly repeated strings of numbers in the database. By June 2022, 55 of 546 investigations (10%) had been concluded; 47 concluded with a retraction (n = 31) or expression of concern (n = 16), of which 41 were published, with 6 planned retractions not yet retracted 9 months after the initial decision. In 8 investigations, editors stated there was insufficient evidence of wrongdoing. In 1 case, the journal limited its investigation to the data set that was provided in the review process and did not consider a problematic data set provided by an author who stepped down after publication. In another case, the journal did not assess available original data themselves but relied on assessment by an expert designated by the author. There were large differences between journals in terms of more vs fewer problem articles, shared vs not providing insight in the assessment process, or being more vs less responsive.

**Conclusions** Procedures recommended by COPE for investigating concerns about data integrity are not always effective, and many editors appear not to know how to handle them.

## Reference

1. Committee of Publications Ethics. Homepage. Accessed July 1, 2022. <https://publicationethics.org/>

<sup>1</sup>Department of Obstetrics and Gynecology, Monash University, Clayton, Australia, ben.mol@monash.edu; <sup>2</sup>Aberdeen Centre for Women's Health Research, Institute of Applied Health Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK; <sup>3</sup>Department of Obstetrics and Gynecology, Nottingham University, Nottingham, UK

**Conflict of Interest Disclosures** Ben W. Mol is supported by a National Health and Medical Research Council Investigator grant

(GNT1176437) and has received consultancy fees from ObsEva and Merck and travel support from Merck.

## Assessment of Submission Withdrawals to a Journal in 2020 and 2021

Catherine M. Ketcham,<sup>1</sup> Martha W. Simmons,<sup>2</sup> Gene P. Siegal<sup>2</sup>

**Objective** It is considered unethical for an author to submit the same manuscript to 2 or more journals simultaneously because it places an undue burden on editors and reviewers and could lead to dual publication. The International Committee of Medical Journal Editors recommends against duplicate submission, and more than 5500 journals follow their guidelines. However, some authors do submit to multiple journals simultaneously, but the extent of this practice is unknown. The editors of *Laboratory Investigation (LI)*, a basic and translational pathology research journal, therefore developed a study question for qualitative analysis: Were any manuscripts that were withdrawn from *LI* during peer review and subsequently published elsewhere under review at both journals concurrently?

**Design** PubMed searches were performed using the titles, keywords, and authors for 36 manuscripts withdrawn from *LI* via email from 2020 to 2021. In that time frame, *LI* received 1550 new submissions, and 787 manuscripts were sent to peer review. The submission, revision, acceptance, and publication dates were collected for the withdrawn articles that were published elsewhere and compared with the submission, decision, and withdrawal dates from *LI*.

**Results** Thirty-six email requests to withdraw submissions were received in 2020–2021. The reasons authors gave were categorized as follows: problems with the data (18), inability to revise the manuscript (6), other concerns (5), and no explanation (7). Twenty of the 36 withdrawn manuscripts had been published in other journals as of January 8, 2022. Of these, 17 (85%) had been under consideration at *LI* and the publishing journal at the same time. All of the duplicate submissions were from China. Multiple requests to withdraw had similar language, though they were from different authors. Five emails contained the phrase “we feel that we have not yet studied our work completely and some new great results are discovered.” Four other messages said that “some updates should be added to this manuscript and it should be rearranged.” Four more stated, “My tutor said that there are certain problems in the current experimental content.” The *LI* editorial office also manages a clinical pathology journal, *Modern Pathology*, which had no email withdrawal requests nor known instances of duplicate submission in 2020–2021. It was unknown whether any other journals had problems similar to those experienced by *LI*.

**Conclusions** The Committee on Publication Ethics recommends against punitive actions for duplicate submission and prefers an educational approach, at least in the first instance. However, these results may indicate that the practice is a deliberate strategy rather than a misunderstanding of the scientific publication process.

Perhaps it is time for publishers and editors to adopt broad screening of submitted manuscripts for dual submissions and to reevaluate the way infractions are addressed.

<sup>1</sup>United States and Canadian Academy of Pathology, Palm Springs, CA, USA, catherine@uscap.org; <sup>2</sup>Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA

**Conflict of Interest Disclosures** Catherine M. Ketcham is a paid employee of the United States and Canadian Academy of Pathology (USCAP), the organization that owns the journal that is the subject of this abstract, *Laboratory Investigation (LI)*. Martha W. Simmons is the senior editorial assistant for *LI* and receives payment from USCAP. Gene P. Siegal is the editor in chief of *LI* and receives a stipend from USCAP.

## Perspectives on Early Warning Signs of Research Fraud or Misconduct

Lisa Parker,<sup>1,2</sup> Stephanie Boughton,<sup>3</sup> Rosa Lawrence,<sup>2</sup> Lisa Bero<sup>2</sup>

**Objective** There is no validated or empirically informed agreement on early warning signs of research fraud or misconduct. This study aimed to explore definitions, early warning signs, and identification strategies for research fraud and misconduct to inform the future development of a screening tool and wider efforts to reduce research fraud and misconduct.

**Design** Semistructured interviews were conducted with experts in systematic reviews, biomedical publishing, and identifying/preventing research fraud and misconduct. Participants were identified through snowball sampling and were recruited via email. Participants were asked about their experiences with research fraud and misconduct, how they define and identify potentially problematic studies, and what they would recommend for inclusion in a screening tool. A thematic analysis approach was used to identify major concepts. This study was approved by the University of Sydney Human Research Ethics Committee. The COREQ reporting guidelines were followed.

**Results** Forty-nine potential participants were contacted and 30 were interviewed. (Reasons for not interviewing included nonresponse [ $n = 14$ ], not available [ $n = 4$ ], passed invitation to colleagues [ $n = 1$ ], and email failure [ $n = 1$ ]. Contacts could have more than 1 reason for not interviewing.) Participants were from 12 countries, including 4 low- and middle-income countries, and had expertise or experience in meta-research ( $n = 12$ ), research ( $n = 11$ ), publishing ( $n = 8$ ), and whistleblowing ( $n = 8$ ). Participants described research fraud as a growing issue, with a lack of widely accessible resources or education to assist in flagging problematic studies. They discussed a range of early warning signs that could be contained in a screening tool for use at either the prepublication or postpublication stage. Signs included no ethics/protocol registration; alerts on existing platforms (eg, Retraction Watch, PubPeer, plagiarism detectors, Stats Check); submission irregularities (eg, unexplained authorship changes, peer review concerns); design issues (eg, aim not meaningful, method implausible); and problematic results

(eg, figure manipulation, extreme outlying data, improbable numbers in randomized clinical trial baseline data). Participants expressed concerns about pressure on researchers to publish, which facilitated high-volume, low-quality outputs. There was discussion about a range of upstream changes to the academic and publishing systems that might reduce production and/or publication of problematic studies, including research fraud.

**Conclusions** Participants described many early warning signs of research fraud or misconduct. Collating these into a screening tool may be useful for reviewers, editors, and publishers. An open-access collection of resources with detailed information on further investigation of potentially problematic studies could help research educators to develop and disseminate knowledge and skills in these areas. Longer-term, upstream changes are also needed to enhance research quality and reduce misinformation.

<sup>1</sup>School of Pharmacy, Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia, lisa.parker@sydney.edu.au; <sup>2</sup>Center for Bioethics and Humanities, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; <sup>3</sup>Cochrane UK, London, UK

**Conflict of Interest Disclosures** Stephanie Boughton is employed by Cochrane as research integrity editor. Lisa Bero is senior research integrity editor, Cochrane, for which the University of Colorado receives remuneration. Lisa Bero is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

**Additional Information** Lisa Bero is a co-corresponding author.

## Pandemic Science

### Characteristics of COVID-19 Clinical Trial Preprints and Associated Publications

Jennifer Klavens,<sup>1</sup> Emily Inwards,<sup>1</sup> Amanda C. Adams,<sup>2</sup> Brian W. Roberts,<sup>1</sup> Timothy F. Platts-Mills,<sup>3</sup> Christopher W. Jones<sup>1</sup>

**Objective** To describe study features and publication characteristics of COVID-19-related clinical trial preprints.

**Design** This was a cross-sectional analysis of COVID-19-related clinical trials published as preprints. Preprints were included if they were uploaded between January 1 and December 31, 2020, to any open-access preprint server indexed by the National Institutes of Health iSearch COVID-19 portfolio and described results from a clinical trial assessing an intervention related to the treatment or prevention of COVID-19. A single investigator assessed manuscripts for eligibility by reviewing the titles and abstracts for all preprints within the iSearch registry, followed by a full-text review of potentially eligible preprints. MEDLINE, Google Scholar, and Embase were then searched to identify peer-reviewed publications matching the included preprints. Two investigators, including a medical librarian, independently searched for published articles at least 14 months after the initial preprint posting date. Two investigators abstracted information from each eligible

preprint, including basic study characteristics (location, funding, size, intervention type), features affecting risk of bias (prospective registration, allocation method, blinding), and trial outcome. Descriptive statistics are reported for these study features, and Cox proportional hazards regression was used to determine associations between key study characteristics and peer-reviewed publication status.

**Results** A total of 22,615 preprints were screened for eligibility and 145 met inclusion criteria. These included preprints uploaded to 5 different preprint servers (medRxiv, Research Square, SSRN, preprints.org, and bioRxiv). Funding sources included government for 65 (45%), industry for 32 (22%), a university or health system for 28 (19%), and a foundation for 16 (11%). The most commonly assessed interventions were drugs (86 [59%]), convalescent plasma (19 [13%]), and vaccines (16 [11%]). Median enrollment was 78 participants (IQR, 30-174 participants). Most trials were registered (136 [94%]), although less than half were registered prospectively (71 [49%]). Blinding was reported for participants in 35 trials (24%), investigators and research staff in 27 (19%), and outcome assessors in 34 (23%). Among 106 trials with more than 1 treatment group, 89 (84%) randomly allocated participants between treatments. Matching peer-reviewed publications were identified for 118 of 145 preprints (81%) in 77 different journals. The median time elapsed between preprint posting and peer-reviewed publication was 115 days (IQR, 55-201 days). Cox proportional hazards analysis revealed significant associations with peer-reviewed publication for government funding and study size (**Table 67**).

**Conclusions** Most COVID-19 clinical trial preprints had undergone peer-reviewed publication, although delays of several months were common. A high proportion were not prospectively registered. Significant associations were not observed between peer-reviewed publication and study characteristics related to risk of bias.

<sup>1</sup>Department of Emergency Medicine, Cooper Medical School of Rowan University, Camden, NJ, USA, jones-christopher@cooperhealth.edu; <sup>2</sup>Medical Library, Cooper Medical School of Rowan University, Camden, NJ, USA; <sup>3</sup>Ophirex Inc, Corte Madera, CA, USA

**Table 67. Multivariable Cox Proportional Hazards Regression Model Describing Association Between Study Characteristics and Peer-Reviewed Publication**

Trial characteristic	Hazard ratio (95% CI) <sup>a</sup>	P value
Government funding	1.58 (1.06-2.36)	.02
Industry funding	1.37 (0.83-2.25)	.20
Foundation funding	1.44 (0.77-2.59)	.27
Prospective registration	1.13 (0.73-1.75)	.58
Randomized treatment allocation	1.28 (0.80-2.04)	.31
Blinded participants	0.51 (0.13-1.97)	.33
Blinded outcome assessors	3.89 (1.00-15.17)	.05
Number of participants	1.02 (1.00-1.03)	.03

<sup>a</sup>A hazard ratio greater than 1 indicates increased likelihood of publication.

**Conflict of Interest Disclosures** Timothy F. Platts-Mills is an employee of Ophirex. Christopher W. Jones reported receiving grants from AstraZeneca, Abbott, Vapotherm, and Ophirex.

**Funding/Support** This work was supported by US Department of Health and Human Services Office of Research Integrity grant ORIIR200066-01-00.

**Role of the Funder/Sponsor** The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the abstract for presentation.

**Disclaimer** The contents of this abstract are those of the authors and do not represent the official views of or an endorsement by the Office of the Assistant Secretary for Health, US Department of Health and Human Services, or the US government.

## Results Availability and Timeliness of Registered COVID-19 Clinical Trials During the First 18 Months of the Pandemic

Maia Salholz-Hillel,<sup>1</sup> Nicholas J. DeVito<sup>2</sup>

**Objective** Transparent, timely results dissemination prevents bias and reduces waste in clinical research. Global emergencies, like COVID-19, increase the necessity of timely and complete reporting.<sup>1</sup> The World Health Organization (WHO) has emphasized the importance of rapid sharing of clinical data and results during public health emergencies.<sup>2</sup> The Dissemination of Registered COVID-19 Clinical Trials (DIRECCT) project examines how and when COVID-19 clinical trial results are disseminated.

**Design** Trials completed during the first 18 months of the pandemic on interventions for the treatment and prevention of acute COVID-19 were examined using the WHO International Clinical Trials Registry Platform COVID-19 study database. Following a minimum delay of 6 weeks between trial completion and searches, automated and manual searches for results publications in PubMed were conducted using Peer Review of Electronic Search Strategies (PRESS) for COVID-19 trial publications from the COVID-evidence project and were supplemented by automated searches for trial identification numbers in the CORD-19 database, manual checks of trial registries and the Cochrane COVID-19 Study Register, and manual keyword searches in trial identification numbers and targeted keywords across CORD-19, the Cochrane COVID-19 Study Register, PubMed, Europe PMC, Google Scholar, and Google. Reporting rates overall and by dissemination route (ie, journal article, preprint, or registry) were described, and time to report across routes was evaluated using survival analysis methods. In addition, data from the registered intervention arms from all trials were extracted to characterize reporting of the most commonly studied COVID-19 interventions.

**Results** Overall, 2621 trials completed between January 1, 2020, and June 30, 2021, were included in the analysis population and searched for results; 1638 (62%) were searched by 1 author. In the preliminary analysis of the first 6 months of the pandemic (285 trials, all dual searched), 41

trials (14%) had results available by August 15, 2020.<sup>3</sup> The most common dissemination route was preprints (25 trials) followed by journal articles (18) and registry results (2); of these, only 4 trials were available as both a preprint and a journal article. The cumulative incidence of any reporting surpassed 20% at 119 days from completion.

**Conclusions** COVID-19 trials completed during the first 6 months of the pandemic did not consistently yield rapid results in the literature or on registries; however, preprints played an important role in results dissemination. These preliminary findings suggest results may be appearing more rapidly compared with clinical trial publication practices prior to the pandemic. Considering trials completed during the first 18 months of the pandemic will offer a more comprehensive picture of trial reporting during COVID-19. The variable quality of registry data potentially limits the precision and completeness of these analyses.

## References

1. Jones CW, Adams AC, Murphy E, et al. Delays in reporting and publishing trial results during pandemics: cross sectional analysis of 2009 H1N1, 2014 Ebola, and 2016 Zika clinical trials. *BMC Med Res Methodol.* 2021;21:120. doi:10.1186/s12874-021-01324-8
2. World Health Organization. Developing global norms for sharing data and results during public health emergencies: statement arising from a WHO consultation held on 1-2 September 2015. Accessed June 20, 2022. [http://www.who.int/medicines/ebola-treatment/blueprint\\_phe\\_datashare-results/en/](http://www.who.int/medicines/ebola-treatment/blueprint_phe_datashare-results/en/)
3. Salholz-Hillel M, Grabitz P, Pugh-Jones M, Streh D, DeVito NJ. Results availability and timeliness of registered COVID-19 clinical trials: interim cross-sectional results from the DIRECCT study. *BMJ Open.* 2021;11(11):e053096. <https://doi.org/10.1136/bmjopen-2021-053096>

<sup>1</sup>QUEST Center for Responsible Research, Berlin Institute of Health at Charité Universitätsmedizin Berlin, Berlin, Germany, maia.salholz-hillel@bih-charite.de; <sup>2</sup>Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

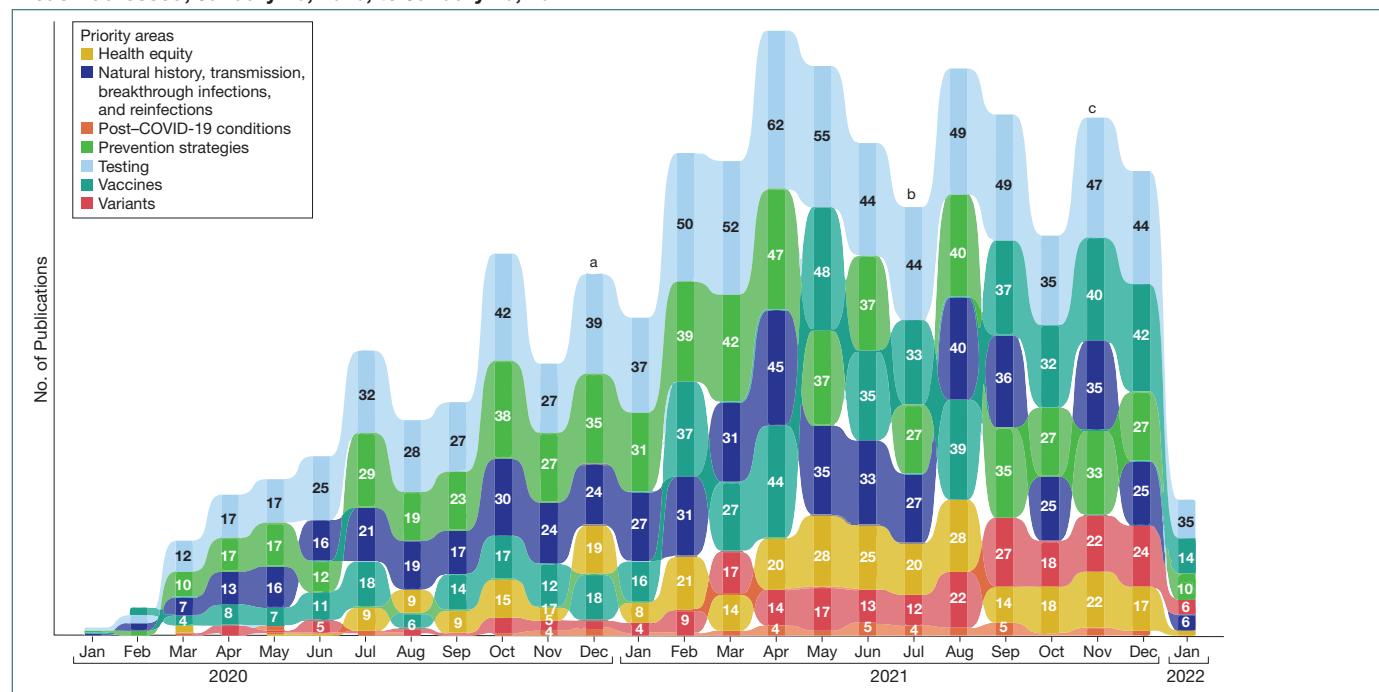
**Conflict of Interest Disclosures** Maia Salholz-Hillel is employed as a researcher under grants from the German Bundesministerium für Bildung und Forschung (BMBF). Outside the submitted work, Nicholas J. DeVito has received a doctoral studentship from the Naji Foundation, has received grant support from the FetzerFranklin Memorial Fund, and has been employed under grants from the Laura and John Arnold Foundation and Good Thinking Society. No other disclosures were reported.

**Funding/Support** This project received funding support through grant FKZ:01KX2021 from the BMBF.

**Role of Funder/Sponsor** The funders had no involvement in the study design, analysis, or writing of the abstract nor the decision to submit.

**Additional Contributions** We acknowledge our colleagues on the Dissemination of Registered COVID-19 Clinical Trials (DIRECCT) research team who contributed to this project, primarily with extensive manual searches: Peter Grabitz, Nicole Hildebrand, Molly

**Figure 26. COVID-19 Publications From the Centers for Disease Control and Prevention (N = 1044) by Public Health Science Priority Areas Addressed, January 20, 2020, to January 20, 2022**



<sup>a</sup>CDC-19 vaccine approval, December 2020.

<sup>b</sup>Emergence of Delta variant, July 2021.

<sup>c</sup>Emergence of Omicron variant, November 2021.

Pugh-Jones, Susanne Schorr, Tjada Schult, Johannes Schweitering, and Daniel Streh.

### COVID-19 Public Health Scientific Publications From the Centers for Disease Control and Prevention, January 2020 to January 2022

Elissa Meites,<sup>1</sup> Martha Knuth,<sup>1</sup> Kaely Hall,<sup>1</sup> Elizabeth Stephenson,<sup>1</sup> Patrick Dawson,<sup>1</sup> Teresa W. Wang,<sup>1</sup> Wei Yu,<sup>1</sup> Muin Khoury,<sup>1</sup> Barbara Ellis,<sup>1</sup> Brian A. King<sup>1</sup>

**Objective** High-quality scientific evidence is critical to support public health decision-making. During public health emergencies, including the COVID-19 pandemic, the US Centers for Disease Control and Prevention (CDC) has been an important source of scientific information. The CDC can provide timely data on and evidence from fundamental public health activities, including domestic and international epidemiologic investigations, laboratory detection, disease surveillance and research, and other scientific work.<sup>1</sup> Routinely, all manuscripts authored by CDC scientists are reviewed internally for scientific quality before publication through a rigorous multilevel process involving subject matter experts and other CDC staff.<sup>2</sup> The objective of this analysis was to understand the CDC's contributions to new COVID-19 science.

**Design** A bibliometric analysis was conducted of the CDC's COVID-19 public health scientific publications and their effect from January 20, 2020, to January 20, 2022, using a quality improvement approach (SQUIRE, version 2.0). COVID-19 scientific articles were catalogued if they had at least 1 CDC-affiliated author, had been reviewed and cleared internally,

and were subsequently published in a peer-reviewed scientific journal and indexed in the World Health Organization's COVID-19 database. The priority topics addressed were identified according to the CDC's COVID-19 Public Health Science Agenda and were assigned tags using keyword scripts in EndNote, and their scientific effects were assessed using Scopus citations, news media, social media impressions, and Altmetric attention scores. This work received a nonresearch determination from the CDC.

**Results** During the first 2 years of the pandemic, CDC authors contributed to 1044 unique COVID-19 scientific publications in 208 journals. The topic areas of focus commonly included testing (853 publications [82%]); prevention strategies (658 publications [63%]); natural history, transmission, breakthrough infections, and reinfections (587 publications [56%]); vaccines (567 publications [54%]); health equity (308 publications [30%]); variants (232 publications [22%]); and post–COVID-19 conditions (44 publications [4%]) (Figure 26). In addition, the CDC adapted to changing key scientific questions as the pandemic evolved, addressing issues including COVID-19 testing expansions; COVID-19 vaccine effectiveness, safety, duration of protection, and access; emergence of new viral variants of concern; and improving health equity.<sup>3</sup> These publications were cited 40,427 times in the scientific literature, have been the topic of 81,921 news media reports and 1,058,893 social media impressions, and received a combined total Altmetric attention score of 920,763. Preprints were not included in this analysis.

**Conclusions** Throughout the COVID-19 pandemic, the CDC has supported the development of high-quality science oriented to improving public health outcomes. The lessons learned from this evaluation included: (1) the agency's COVID-19 Public Health Science Agenda helped guide scientific activities; (2) scientific manuscripts developed, reviewed, and published to address priority topics became highly effective; and (3) the CDC is committed to monitoring emerging issues and addressing gaps in evidence needed to improve health outcomes. Data-driven strategies are essential to address disparities and improve the health outcomes of people disproportionately affected by COVID-19.

## References

1. Bunnell R, Ryan J, Kent C; CDC Office of Science, CDC Excellence in Science Committee. Toward a new strategic public health science for policy, practice, impact, and health equity. *Am J Public Health*. 2021;111(8):1489-1496. doi:10.2105/AJPH.2021.306355
2. Cono J, Jaffe H. The CDC clearance process: supporting quality science. *Am J Public Health*. 2015;105(6):e1-e2. doi:10.2105/AJPH.2015.302691
3. Cohn AC, Mahon BE, Walensky RP. One year of COVID-19 vaccines: a shot of hope, a dose of reality. *JAMA*. 2022;327(2):119-120. doi:10.1001/jama.2021.23962

<sup>1</sup>Chief Science Office, COVID-19 Response, Centers for Disease Control and Prevention, Atlanta, GA, USA, emeites@cdc.gov

**Conflict of Interest Disclosures** None reported.

**Funder/Sponsor** The Centers for Disease Control and Prevention (CDC).

**Role of the Funder/Sponsor** The CDC participated in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Disclaimer** The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

**Additional Information** Martha Knuth is a co-corresponding author.

## Brazilian Researchers and Journal Editors Experiences With Scientific Publication During the COVID-19 Pandemic

Luísa von Zuben Veçoso,<sup>1</sup> Marcus Tolentino Silva,<sup>2</sup> Taís Freire Galvão<sup>1</sup>

**Objective** To assess whether strategies adopted by Brazilian journals for COVID-19 articles are associated with the increase of article submission and publication in 2020. Early assessments of medical journals worldwide since the outbreak of the pandemic showed that COVID-19-related articles were published faster than all articles before the pandemic.<sup>1</sup>

**Design** An online survey was conducted with Brazilian researchers from October to December 2020. After ethical approval, a questionnaire with closed questions about the

associations of the pandemic with editorial tasks was sent by email. This analysis included the questionnaires answered by editors in chief who identified their journal name to assess whether policies adopted by Brazilian journals were associated with outcomes for submissions and publications in the first semester of 2020. Information on journal subject area, index source, open access status, whether the publisher was public or private, and charges for submission or publication was collected from journal websites. The 2021 journal impact factor (JIF) was also collected for all journals. Adjusted odds ratio (ORs) and 95% CIs were calculated by logistic regression adjusted by private or public publisher, charging status, and JIF.

**Results** In total, 1299 Brazilian researchers responded to the survey. Of 163 editors in chief who participated, 87 (53%) provided their journal name and were included in the analysis. Journal subject areas included the humanities (n = 43), life sciences (n = 41), and exact sciences (n = 3). Most journals were open access (n = 85), one-third were indexed in Web of Science (n = 29), and a 2021 JIF was available for 18. Thirty-two publishers were private, and 21 charged for submission or publication. The scope of 58 journals covered topics related to COVID-19. Journals that adopted fast-track publishing had a greater increase in manuscript submissions (OR, 3.85; 95% CI, 1.42-10.41) and publications (OR, 3.03; 95% CI, 1.20-7.63). Journals that invited manuscripts about COVID-19 also had a greater increase in submissions (OR, 2.89; 95% CI, 1.13-7.40) and publications (OR, 2.57; 95% CI, 1.03-6.38). The increase in publications was greater in journals that started to accept articles published as preprints (OR, 4.26; 95% CI, 1.26-14.38) and that opened a call for COVID-19 manuscripts (OR, 3.49; 95% CI, 1.22-10.00), whereas those that published an editorial about COVID-19 had a greater increase in submissions (OR, 2.65; 95% CI, 1.02-6.88). Publishing a dedicated COVID-19 issue did not affect the outcomes (**Table 68**).

**Conclusions** Editorial changes in response to the COVID-19 pandemic's challenges were associated with an increase in the

**Table 68. Frequency of Adopted Strategies and ORs for Increases in Article Submissions and Publications by 87 Journals in Brazil in Response to the COVID-19 Pandemic**

Strategy in response to pandemic	No.	OR (95% CI)	
		Increase in submissions	Increase in publications
Started to accept submission of studies published as preprint	15	3.88 (1.00-15.01)	4.26 (1.26-14.38)
Opened a call for manuscripts about COVID-19	22	2.89 (0.96-8.64)	3.49 (1.22-10.00)
Adopted fast tracking for COVID-19 submissions	32	3.85 (1.42-10.41)	3.03 (1.20-7.63)
Published editorial on COVID-19	33	2.65 (1.02-6.88)	1.74 (0.70-4.33)
Published or planned to publish invited articles on COVID-19	32	2.89 (1.13-7.40)	2.57 (1.03-6.38)
Published or planned to publish a dedicated issue for COVID-19	21	2.50 (0.81-7.75)	1.55 (0.53-4.51)

Abbreviation: OR, odds ratio.

number of submissions and publications by Brazilian journals. These results are limited by the small size and nonrepresentativeness of the sample.

## Reference

1. Horbach SPJM. Pandemic publishing: medical journals strongly speed up their publication process for COVID-19. *Quant Sci Stud.* 2020;1(3):1056-1067. doi:10.1162/qss\_a\_00076

<sup>1</sup>Faculty of Pharmaceutical Sciences, State University of Campinas, Campinas, São Paulo, Brazil, luisavecoso@gmail.com; <sup>2</sup>Graduate Program of Pharmaceutical Sciences, University of Sorocaba, Sorocaba, São Paulo, Brazil

**Conflict of Interest Disclosures** Luísa von Zuben Veçoso is an employee of Boehringer Ingelheim in Brazil. The company has no connection to this work.

**Additional Information** Taís Freire Galvão is a co-corresponding author.

## Peer Review

### Unprofessional Comments in Peer Review Reports Across Scholarly Disciplines

Mario Malički,<sup>1</sup> Taym Alsalti,<sup>2</sup> Daniel García-Costa,<sup>3</sup> Francisco Grimaldo,<sup>3</sup> Elena Álvarez-García,<sup>3</sup> Ana Jerončić,<sup>4</sup> Steven M. Goodman,<sup>5</sup> Flaminio Squazzoni,<sup>6</sup> Bahar Mehmani<sup>7</sup>

**Objective** Previous research has indicated that 642 of 1116 surveyed researchers (58%) received unprofessional comments in peer review reports at least once in their professional life<sup>1</sup> and that 179 of 1491 review reports (12%) in the fields of ecology and evolution and behavioral medicine contained unprofessional comments.<sup>2</sup> It was the goal of this study to estimate the percentage of manuscripts with at least 1 unprofessional comment made in a broad sample of peer review reports across disciplines.

**Design** This was a cross-sectional study of review reports available in the PEERE database, which covers 10 years of peer review records across all impact factor quartiles and peer review types and includes journals from the fields of life sciences, health and medicine, physical sciences, and social sciences and economics.<sup>3</sup> Sample size calculation indicated that 380 of 297,026 manuscripts contained in the PEERE database should be analyzed to detect a prevalence of 1% of manuscripts (range, 0%-2%) having at least 1 unprofessional comment. Randomized stratified sampling was used to preserve scholarly field distribution. All review reports of 1147 sampled manuscripts were then extracted. Reading of the reports was done independently by 3 researchers, who marked instances of unprofessional comments. Disagreement between researchers was resolved through consensus.

**Results** Of 1147 analyzed review reports, 13 (1.1%) contained at least 1 unprofessional comment (**Table 69**). All were found in review reports of different manuscripts; authors of 13 of 380 analyzed manuscripts (3.4%) received unprofessional comment(s) during peer review. All instances

**Table 69. Unprofessional Comments Found in Review Reports**

Comment	Discipline
It seems like the article has not been accurately reviewed by an experienced researcher.	Health and medical sciences
First they need to learn how to properly structure a biomedical research paper.	Health and medical sciences
I do not think this is a shortcoming in my ability to understand the material, but an inability of the authors to clearly and appropriately explain their work...This raises further concern about the authors' grasp of the subject material, particularly regarding wildlife ecology and the existing literature on corridor studies. ...I am concerned about the authors' understanding of the subject matter, particularly dealing with aspects of wildlife ecology and landscape use.	Life sciences
Statement like these "E2* can transfer energy to the dissolved oxygen to generate ROS such as OH and O2 in the solution and thereby promote the degradation of E2" makes one wonder whether the authors understand the basic premise of photochemistry.	Physical sciences
If the authors had familiarized themselves with the ample work in modeling these types of systems, they would not have given this title to this section.	Physical sciences
But in what century do these authors live?	Physical sciences
I cannot imagine that there is still researcher doing such a simple work and submit it to [anonymized] journal.... It seems that the authors know nothing about the state-of-the-art works in denoising.	Physical sciences
The study design is, to put it mildly, rather naive.	Physical sciences
The authors had better read some relevant papers.	Physical sciences
SBSE is not the focus of your paper this is the wrong term in the wrong place that prove a sloppy attitude, plus, who is the judge?	Physical sciences
The list of flaws is so large that cannot be included here.... This is also combined with a worrying lack of attention to details.	Physical sciences
This paper more looks like a masters thesis and its most of the materials can be found in any preliminary statistics text book... Moreover, it is very irritating to find all the ACF, PACF and CCF plots in the paper, which show the immaturity of the authors.	Physical sciences
Are the authors naive enough to assume that all forms of design have the same technical and social requirements and patterns?	Physical sciences

of unprofessional comments were found among detailed review reports.

**Conclusions** In the study sample, 3.4% of submitted manuscripts received at least 1 unprofessional comment during peer review. Although the PEERE database presents the largest collection of confidential review reports shared by publishers, it might not be representative of all journals. Nevertheless, to the best of our knowledge, this is the first study of unprofessional comments that used random sampling across scholarly disciplines. Future studies should evaluate the effects of these comments on the authors, assess whether and how they respond to them, and explore automated approaches for detecting unprofessional comments.

## References

1. Silbiger NJ, Stubler AD. Unprofessional peer reviews disproportionately harm underrepresented groups in STEM. *Peer J.* 2019;7:e8247. doi:10.7717/peerj.8247

2. Gerwing TG, Gerwing AMA, Avery-Gomm S, Choi CY, Clements JC, Rash JA. Quantifying professionalism in peer review. *Res Integ Peer Rev*. 2020;5(1):1-8. doi:10.1186/s41073-020-00096-x

3. Squazzoni F, Bravo G, Farjam M, Marusic A, Mehmani B, Willis M, et al. Peer review and gender bias: a study on 145 scholarly journals. *Sci Advanc*. 2021;7(2):eabdo299. doi:10.1126/sciadv.abdo299

<sup>1</sup>Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, USA, mario.malicki@mefst.hr; <sup>2</sup>Personality Psychology and Psychological Diagnostics, University of Leipzig, Leipzig, Germany; <sup>3</sup>Department of Computer Science, University of Valencia, Burjassot-Valecia, Spain; <sup>4</sup>Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia; <sup>5</sup>Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA; <sup>6</sup>Department of Social and Political Sciences, University of Milan, Milan, Italy; <sup>7</sup>STM Journals, Elsevier, Amsterdam, the Netherlands

**Conflict of Interest Disclosures** Mario Malički is the co-editor in chief of *Research Integrity and Peer Review*. Bahar Mehmani is reviewer experience lead at Elsevier. Other authors declare no competing interests. Steven Goodman and Bahar Mehmani are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract.

**Funding/Support** The work by Francisco Grimaldo, Daniel García-Costa, and Elena Álvarez-García is supported by the Spanish Ministry of Science, Innovation and Universities, the Spanish State Research Agency, and the European Regional Development Fund under project RTI2018-095820-B-I00.

**Role of Funder/Sponsor** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

### Preference and Characteristics of US-Based Authors for Single- vs Double-Anonymous Peer Review

Meredith S. Campbell Joseph,<sup>1</sup> Amy L. Davidow,<sup>2</sup> Lewis R. First,<sup>3</sup> Alex R. Kemper<sup>4</sup>

**Objective** In May 2020, *Pediatrics* moved to a new author submission platform that allows authors to choose single-anonymous (SA) or double-anonymous (DA) peer review.<sup>1</sup> Prior to May 2020, SA reviewing was the only option. The purpose of this study was to describe author characteristics associated with choosing SA vs DA peer review. We tested 3 hypotheses: (1) corresponding authors who are women are more likely to select DA peer review than those who are men, (2) corresponding authors who are junior faculty (ie, assistant professors, instructors, or trainees) are more likely to select DA peer review than faculty at higher academic ranks (ie, associate professor or professors), and (3) corresponding authors who select DA peer review are more likely to have their manuscript rejected than those who select SA peer review given that knowledge of a well-respected author by the reviewer may preferentially bias the reviewer favorably.

**Design** In this cross-sectional study, we classified *Pediatrics* articles (submitted between May 4, 2020, and April 1, 2021)

by peer review type and then randomly sampled 150 articles of each type. After excluding sampled articles that were not “regular studies” and others without a US-based author, 169 regular research articles (73 SA and 96 DA peer review) remained. Corresponding author gender and academic rank were determined manually using an internet search. We tested our hypotheses using  $\chi^2$  tests and  $\chi^2$  tests for trend.

**Results** Of the 2720 regular articles submitted to the journal during the study period, 505 (18.6%) were submitted for DA peer review. We found no difference in the proportion of corresponding authors who chose DA peer review by gender (62% men vs 53% women;  $P = .24$ ). There was no statistically significant difference in the likelihood of choosing DA peer review with increasingly higher academic rank ( $\chi^2$  test for trend,  $P = .20$ ). The likelihood of rejection was somewhat higher for DA vs SA peer review; however, this difference was not significant (94.8% vs 86.3%;  $P = .06$ ).

**Conclusions** US-based author preference for DA peer review in *Pediatrics* was not associated with gender or faculty rank. Given that nearly 1 in 5 authors submitting to *Pediatrics* preferred DA peer review, both options will continue to be offered.

### Reference

1. Morrison JM, First LR, Kemper AR. Recommendations for blinded peer review: a survey of high-quality *Pediatrics* reviewers. *Pediatrics*. 2020;146(2):e20201403. doi:0.1542/peds.2020-1403

<sup>1</sup>Division of Neonatology, Department of Pediatrics, Virginia Commonwealth University, Richmond, VA, USA, meredith.campbell@vcuhealth.org; <sup>2</sup>Department of Biostatistics, New York University School of Global Public Health, New York, NY, USA; <sup>3</sup>Department of Pediatrics, The Larner College of Medicine, The University of Vermont and The University of Vermont Children’s Hospital, Burlington, VT, USA; <sup>4</sup>Division of Primary Care Pediatrics, Nationwide Children’s Hospital, Columbus, OH, USA

**Conflict of Interest Disclosures** None reported.

### Peer Review Process and Models

#### Differences in the Style and Quantity of Reviewer Comments in Structured vs Unstructured Peer Review Forms

Emma Ghazaryan,<sup>1</sup> Marina Broitman,<sup>2</sup> Harold Sox<sup>2</sup>

**Objective** This study compared 2 critique formats used for subject matter expert (SME) reviews for Patient-Centered Outcomes Research Institute (PCORI) draft final research reports (DFRRs).<sup>1</sup> In November 2018, PCORI changed SME critique formats from 13-item structured forms to unstructured forms after DFRR authors commented about the burden and repetitiveness of critiques. Existing review quality tools were examined<sup>2,3</sup> but focused more on scientific review quality than comment type.

**Design** Subject matter expert reviews from all clinical research reports that completed peer review between October

2017 and September 2019 (13 months before and 11 months after the review form change) were compared to see whether structured and unstructured reviews had the same number and types of comments. In September 2019, the abstract authors defined comment categories for SME reviews: (1) general positive or complimentary comments, (2) critical comments leading to a report change, (3) critical comments not leading to a change, (4) grammatical errors, (5) comments repeating an earlier statement by the same reviewer, and (6) neutral factual comments. The first author (E.G.) developed the categories based on previous tools<sup>a</sup> and iteratively revised the categories on the basis of discussions and coding comparisons with the second author (M.B.). The responses of DFRR authors to the SME review comments were used to confirm types of comments. The study outcome was the number of comments in each category as rated by the first author. Interrater reliability was determined by category agreement between the first author and a colleague unfamiliar with this project on 124 reviewer comments, resulting in a kappa of 0.69. The percentage of each comment category and the mean (SD) number of comments per SME review were calculated. Given the small sample, analyses were descriptive.

**Results** The sample included 49 DFRRs (30 before [61%] and 19 after [39%] the review form change), with a total of 99 SME reviews (2.1 SME reviews per DFRR); 61 (62%) used the structured form and 38 (38%) used the unstructured form. Including all comment categories, the mean (SD) number of comments per review was 39.4 (17.7) with the structured form and 17.7 (9.9) with the open-ended form (**Table 70**). With the unstructured form, positive or complimentary comments were a larger percentage of all comments (36.7% vs 30.4%), whereas the percentage of repetitive comments was lower (2.2% vs 13.4%) compared with the structured form. With both forms, approximately 40% of comments were critical comments that led to a change in the DFRR.

**Conclusions** Structured review forms provided more reviewer comments than unstructured forms but included more repetitive comments and a smaller proportion of positive comments. The study sample was too small to address the association of these differences with the quality or speed of peer review.

## References

1. Broitman M, Sox HC, Slutsky J. A model for public access to trustworthy and comprehensive reporting of research. *JAMA*. 2019;321(15):1453-1454. doi:10.1001/jama.2019.2807
2. van Rooyen S, Black N, Godlee F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. *J Clin Epidemiol*. 1999;52(7):625-629. doi:10.1016/s0895-4356(99)00047-5
3. Superchi C, González JA, Solà I, Cobo E, Hren D, Boutron I. Tools used to assess the quality of peer review reports: a methodological systematic review. *BMC Med Res Methodol*. 2019;19(1):48. doi:10.1186/s12874-019-0688-x

**Table 70. Number and Percentage of Comments per Category per Review Critique**

Comment category	Structured reviews (n = 61)		Open reviews (n = 38)	
	No. of comments per review, mean (SD)	Mean total comments per review, %	No. of comments per review, mean (SD)	Mean total comments per review, %
Total comments per review	39.4 (17.7)	100	17.7 (9.9)	100
Positive comments	12.0 (6.6)	30.4	6.5 (4.1)	36.7
Led to a change	15.9 (10.9)	40.3	7.0 (6.2)	39.5
Did not lead to a change	3.0 (2.9)	7.6	1.2 (1.5)	6.7
Comments on spelling and grammar	0.8 (1.4)	2	1.0 (2.7)	5.6
Repetitive comments	5.3 (5.7)	13.4	0.4 (0.7)	2.2
Neutral comments	2.4 (5.6)	6.1	1.7 (3.5)	9.6

<sup>a</sup>World Health Organization, Yerevan, Armenia; <sup>b</sup>Patient-Centered Outcomes Research Institute, Washington, DC, USA, mbroitman@pcori.org

**Conflict of Interest Disclosures** Emma Ghazaryan's spouse is employed by the National Institutes of Health. Marina Broitman's spouse is employed by the National Institutes of Health. No other disclosures were reported.

**Funding/Support** This work was funded by the Patient-Centered Outcomes Research Institute (PCORI).

**Role of the Funder/Sponsor** PCORI provided review and approval of the abstract for submission to the Peer Review Congress.

**Disclaimer** This work was completed by Emma Ghazaryan when she was employed as an intern at PCORI. Marina Broitman and Harold Sox are PCORI employees. All statements, findings, and conclusions in this publication are solely those of the authors and do not necessarily represent the views of PCORI or its board of governors.

## Open Participation in Open Peer Review: Models, Reviewers, and Concepts

Janaynne Carvalho do Amaral,<sup>1</sup> Eloísa Príncipe<sup>1</sup>

**Objective** One of the changes to the peer review process introduced by open science is the participation of reviewers who were not selected by editors or not indicated by authors in the evaluation of the manuscripts submitted for publication. Previous studies characterizing open peer review and its traits have called this phenomenon “open participation,”<sup>1</sup> “crowdsourced peer review,”<sup>2</sup> and “public peer review.”<sup>3</sup> This study aimed to identify and describe from the literature what peers, concepts, and models of peer review are associated with open participation in scientific journals.

**Design** An integrative review was conducted. The search strategy was used with Scopus, Web of Science, and PubMed. The literature was retrieved in March 2021 and was limited to the languages English, Portuguese, and Spanish but with no time limitations. All data were collected from the final sample. Models of open peer review with open participation were considered as those that contemplated the participation

**Table 71. Peers, Concepts, and Models of Open Peer Review Related to Open Participation Retrieved From the Literature<sup>a</sup>**

Scientific journal(s) or prototype model	Open peer review model	Open peer review characteristics adopted	Peers
<i>The Medical Journal of Australia</i>	It combines traditional peer review with open peer review. Papers approved in traditional peer review go to open peer review if reviewers and authors agree. Readers may comment on the papers, and authors are encouraged to respond to the comments. It publishes the readers' comments together with papers.	Open participation, open identities (optional), open reports, open interaction	Selected reviewers and readers
<i>Atmospheric Chemistry and Physics; Hydrology and Earth System Sciences; Earth System Science; Geoscientific Model Development</i>	Interactive Open Access Peer Review Model composed of 2 stages combined with interactive public discussion: First stage: manuscripts pass for a rapid prescreening and are published in the journal's discussion forum to interactive public discussion for a period of 8 wk; Second stage: peer review and manuscript revision are finalized as in traditional journals and the accepted articles are published.	Open participation, open identities (optional for invited reviewers), open reports, open interaction	Invited reviewers and interested members of the scientific community
Revamped open peer review process <sup>b</sup>	Prototype of an open peer review model with open participation composed of 7 steps: First step: prepublish first then review; Second step: readers and expert reviewers review article; Third step: most reviewed is a determinant of popularity; Fourth step: popularity is a determinant of cost of article; Fifth step: author revises after receiving comments; Sixth step: expert reviewers' names are made public; and Seventh step: reader reviewers' names are made public.	Open participation, open identities, open interaction	Expert reviewers and readers
<i>World Economic Review; Economic Thought: History, Philosophy, and Methodology</i>	Open peer review discussion model composed of 4 steps: First step: publication of the manuscripts in a discussion forum; Second step: receipt and post of comments; Third step: closing of the forum discussion, editorial decision, and publication of the selected comments; and Fourth step: receipt of comments after publication, which can be posted or not.	Open identities (optional), open interaction, open reports (selected by the editor)	Expert reviewers invited by editors and pointed out by authors, members of the World Economics Association (UK)
<i>Journal of Instructional Research</i>	Hybrid review model composed of 2 stages: First stage: open—public review; Second stage: private—peer review.	First stage: open identities (mandatory), open interaction, platforms for commentaries (mandatory registration) Second stage: blind review, reviews closed to the editor	Expert reviewers, the public
<i>Research Involvement and Engagement; The BMJ</i>	It keeps the steps of the traditional peer review but includes new types of reviewers in its evaluation process of manuscripts.	Open identities (mandatory) to academic reviewers, patient reviewers, open reports	Academics, patients, carers, lay persons, the public

<sup>a</sup>The concepts of open peer review models with open participation are open access, interaction, transparency, and crowdsourcing.

<sup>b</sup>Described exactly as reported by the authors Jasni Ahmad and Norshuhada Shiratuddin. This is a prototype proposed for scientific journals; this model still was not implemented.

of the public. These models were characterized according to the terminology of open peer review established by Ross-Hellauer.<sup>1</sup> Self-appointed reviewers were regarded as a new type of reviewer. The concepts around open peer review models with open participation were understood to be the foundation for its implementation.

**Results** A total of 562 studies were retrieved. However, 407 remained after removing duplicates, and 20 met the inclusion criteria. These studies were published in English from 1998 to 2018. Of these 20 studies, 9 (45%) presented implemented models, 7 (35%) discussed public participation and comments, 2 (10%) proposed a model, 1 (5%) approached the quality of the manuscript selection process, and 1 (5%) analyzed the reliability of peer review in open participation. Six open peer review models with open participation were found. These models were totally open or combined with the traditional peer review to discuss manuscripts publicly. These models kept the steps of the traditional peer review process or were divided into 2 or more stages (**Table 71**). The peer reviewers were the readers of the journal, reviewers invited by the editor or author, or patients. The studies that focused on public participation and on comments were questioning the expertise of this audience to evaluate manuscripts, the

incentives to comment, and the potential effect these models may have on the careers of researchers. Public comments may help to improve the quality of the publication and of the manuscript selection process, but they might not increase interrater reliability among reviewers. Open access, crowdsourcing, interaction, and transparency were the main concepts of the models of open peer review with open participation.

**Conclusions** Open peer review models with open participation have different degrees of openness, are expanding the idea of who a peer reviewer can be, and are bringing new challenges to the peer review process. Furthermore, open identities are a sensitive aspect of these models, for which the concepts are aligned with open science values.

## References

- Ross-Hellauer T. What is open peer review? a systematic review. *F1000Res*. 2017;6:588. doi:10.12688/f1000research.11369.2
- Ford E. Defining and characterizing open peer review: a review of the literature. *J Scholarly Publishing*. 2013;44(4):311-326. doi:10.3138/jsp.44-4-001

3. Pöschl U. Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Front Comput Neurosci.* 2012;6:33. doi:10.3389/fncom.2012.00033

<sup>1</sup>Brazilian Institute of Information in Science and Technology/Federal University of Rio de Janeiro, Post-Graduate Program in Information Science, Rio de Janeiro, Brazil, menequete@gmail.com

**Conflict of Interest Disclosures** None reported.

### Peer Reviewers' Willingness to Review and Their Recommendations After the *Finnish Medical Journal* Changed From Single-Anonymous to Double-Anonymous Peer Review

Piitu Parmanne,<sup>1</sup> Joonas Laajava,<sup>2</sup> Noora Järvinen,<sup>3</sup> Terttu Harju,<sup>4,5,6</sup> Mauri Marttunen,<sup>7,8</sup> Pertti Saloheimo<sup>3</sup>

**Objective** Peer reviewers' willingness to review, their recommendations, and quality of reviews were explored after switching from single-anonymous to double-anonymous peer review in the *Finnish Medical Journal*, published in Finnish. Previous research has mainly concerned journals published in English.<sup>1</sup>

**Design** The *Finnish Medical Journal* switched to double-anonymous peer review on September 1, 2017. The material comprised reviews submitted from September 2017 to February 2018. The controls were the reviews submitted between September 2015 and February 2016 and between September 2016 and February 2017. The reviews on all manuscripts with at least 2 reviews were included. In cases with more than 2 reviews for the manuscript, the first 2 reviews received were included. The number of invitations needed to receive 2 reviews for a manuscript was calculated. How often the reviewers recommended to accept as is, minor revision, major revision, or to reject was also explored. The contents of the reviews were independently assessed by 2 experienced reviewers (T.H., M.M.) who were unaware of the peer review model and the decisions made on the manuscripts. The Review Quality Instrument<sup>2</sup> was modified to apply for both original research and review manuscripts. The reviewers' recommendations were tested using the  $\chi^2$  test, and the means of quality assessments were tested with an independent-samples  $t$  test. A  $P$  value  $<.05$  was considered statistically significant.

**Results** A total of 118 reviews given for 59 original research and review manuscripts by 114 individual double-anonymous reviewers were included. These were compared with 232 reviews for 116 manuscripts by 213 reviewers who were single-anonymous, ie, the reviewers were aware of the review process. The number of invitations needed to obtain 2 reviews was similar during the double-anonymous and single-anonymous periods: median, 3 (IQR, 2-4 reviews; range, 2-7 reviews). When performing a double-anonymous review, the reviewers more seldom recommended accept as is and minor revision than during the control period (accept as is, 8.5% vs 12.1%; minor revision, 50.8% vs 60.8%, respectively), and

more often reviewers chose major revision and reject (major revision, 33.1% vs 22.8%; reject, 7.6% vs 4.3%, respectively;  $P = .07$ ). For the quality assessment, 116 reviews were included (in 2 cases there was no written review) and compared with 104 reviews given between September 2016 and February 2017. The results are shown in **Table 72**.

**Conclusions** In general, the quality of double-anonymous reviews was significantly better than that of single-anonymous reviews, contrary to a previous meta-analysis.<sup>1</sup> Switching to double-anonymous review did not alter the reviewer's willingness to review. The reviewers became slightly more critical and more often recommended major revision or reject, but this study was underpowered to show statistical significance. The study period was limited by the change in peer review platform in March 2018. Another limitation was that there was no control for the quality of the manuscripts.

### References

- Bruce R, Chauvin A, Trinquart L, Ravaud P, Boutron I. Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. *BMC Med.* 2016;14(1):85. doi:10.1186/s12916-016-0631-5
- van Rooyen S, Godlee F, Evans S, Black N, Smith R. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ.* 1999;318(7175):23-27. doi:10.1136/bmj.318.7175.23

**Table 72. Quality Assessment of Double-Anonymous vs Single-Anonymous Reviews Using the Modified Review Quality Instrument<sup>a</sup>**

Question	Single-anonymous (n = 104)	Double-anonymous (n = 116)	P value
Did the reviewer discuss the importance of the research question/topic of the review?	3.38 (1.17) (3.22-3.54)	3.53 (1.22) (3.37-3.69)	.19
Did the reviewer discuss the originality of the manuscript?	2.51 (1.34) (2.33-2.69)	2.90 (1.36) (2.72-3.07)	.003
Did the reviewer identify the strengths and weaknesses of the methods/literature search?	2.90 (1.33) (2.71-3.08)	3.13 (1.25) (2.97-3.30)	.06
Did the reviewer make useful comments on writing, organization, tables, and figures?	3.31 (1.10) (3.16-3.46)	3.41 (1.10) (3.27-3.56)	.31
Were the reviewer's comments constructive?	3.61 (0.86) (3.50-3.73)	3.69 (0.80) (3.59-3.79)	.35
Did the reviewer supply appropriate evidence using examples from the manuscript to substantiate their comments?	3.30 (1.22) (3.14-3.47)	3.43 (1.15) (3.28-3.58)	.27
Did the reviewer comment on the authors' interpretation of the results/literature?	3.00 (1.22) (2.83-3.16)	3.21 (1.10) (3.06-3.35)	.06
How would you rate the tone of the review?	3.79 (0.66) (3.70-3.89)	3.77 (0.65) (3.68-3.85)	.69
Mean of assessments on all topics	3.22 (1.20) (3.17-3.28)	3.38 (1.13) (3.33-3.44)	<.001

<sup>a</sup>Values are presented as mean (SD) (95% CI). Mean values are on a 1 to 5 scale, with 5 indicating better and 1 indicating worse.

<sup>1</sup>Finnish Medical Association, Helsinki, Finland; <sup>2</sup>University of Helsinki, Helsinki, Finland; <sup>3</sup>*Finnish Medical Journal*, Helsinki, Finland, pertti.saloheimo@laakarilehti.fi; <sup>4</sup>Respiratory Medicine Research Unit, Department of Medicine, University of Oulu, Oulu, Finland; <sup>5</sup>Medical Research Center, Oulu University Hospital, Oulu, Finland; <sup>6</sup>OYS Somatics, Internal Medicine Centre, Pulmonary Outpatient Clinic, Oulu University Hospital, Oulu, Finland; <sup>7</sup>Department of Adolescent Psychiatry, Helsinki University Hospital and University of Helsinki, Helsinki, Finland; <sup>8</sup>Department of Public Health Solutions, Finnish Institute for Health and Welfare in Finland, Helsinki, Finland

**Conflict of Interest Disclosures** None reported.

**Funding/Support** The study was partially funded by a grant from the Finnish Association for Scholarly Publishing.

**Role of the Funder/Sponsor** The funder had no impact on the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgment** We thank designer Lauri Parikka for technical assistance in collecting the data.

## Preprints

### Adherence to Reporting Guidelines in Systematic Review Preprints and Their Corresponding Journal Publications

Haley K. Holmer,<sup>1</sup> Edi E. Kuhn,<sup>1</sup> Celia V. Fiordalisi,<sup>1</sup> Rose Relevo,<sup>1</sup> Mark Helfand<sup>1</sup>

**Objective** Previous research indicates that adherence to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) checklist is suboptimal, with 9 items adhered to by fewer than 67% of published systematic reviews (SRs).<sup>1</sup> It is unknown whether adherence is similar in the preprint literature and whether adherence is improved in peer-reviewed journal articles. This study compared adherence to reporting standards in preprint SRs and their corresponding journal publications.

**Design** In 50 randomly sampled SR preprints uploaded to *medRxiv* between database inception and December 8, 2021, any preprint SR with at least 1 meta-analysis and a corresponding journal publication was included; the focus of the SR was not considered in the inclusion criteria. The PRISMA 2020 for Abstracts Checklist<sup>2</sup> was used to assess adherence to reporting standards and discrepancies in reporting between preprint-publication pairs. The number of meta-analyses published by each journal in the last 3 years was quantified as a proxy for journal resources to critique meta-analyses. The SRs were classified as adherent if at least 9 PRISMA items (75%) were reported and nonadherent if less than 9 items were reported.

**Results** Of 34,760 preprints on *medRxiv* on December 8, 2021, 922 were SRs, 373 of which were published. Of these, 220 included a meta-analysis, and from these a random sample of 50 preprints was obtained. The included preprints were published in 38 unique journals (median impact factor, 3.5 [IQR, 2.9–4.8]), publishing a median of 34.5 SRs (IQR,

4.5–387.0) with meta-analysis in the last 3 years. Nineteen (38%) were conducted in the US or Canada, and 24 (48%) had registered protocols. Despite 80% of journal publications stating adherence to the PRISMA checklist, 31 (62%) were nonadherent compared with 36 (72%) of the corresponding preprints (odds ratio, 0.63; 95% CI, 0.25–1.59). The items most frequently unreported in preprint-publication pairs were details of included studies (46 [92%]), risk of bias (39 [78%]), and funder (37 [74%]). The mean journal impact factor for nonadherent preprint-publication pairs was similar to the mean journal impact factor for adherent preprint-publication pairs.

**Conclusions** In this sample of SR preprints, adherence to the PRISMA 2020 checklist was low and improved only slightly in corresponding journal publications. Additional analyses examining whether lack of vigilance on the part of journals, journal formatting requirements, priorities of journals and/or authors potentially being out of sync with PRISMA guidelines, or other explanations may account for the lack of improvement with full publication.

## References

1. Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement and extensions: a scoping review. *Syst Rev*. 2017;6(1):263. doi:10.1186/s13643-017-0663-8
2. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71

<sup>1</sup>Portland VA Research Foundation, Portland, OR, USA, haley.holmer@va.gov

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was funded by the Agency for Healthcare Research and Quality Effective Healthcare Program through a contract to the Scientific Resource Center (HHS-A2902017003C).

**Additional Information** The authors of this abstract are responsible for its content. Statements in the abstract do not necessarily represent the official views of or imply endorsement by the Agency for Healthcare Research and Quality, US Department of Health and Human Services.

### Assessment of Manuscripts Submitted to *Annals of Internal Medicine* That Were Posted as Preprints

Jill Jackson,<sup>1</sup> Christine Laine<sup>1</sup>

**Objective** The posting of clinical research reports on preprint servers prior to peer review increased dramatically during the SARS-CoV-2 pandemic.<sup>1</sup> However, little is known about the fate of preprint reports after submission to a peer-reviewed journal. This study investigated *Annals of Internal Medicine*'s experience with submissions available as preprints submitted during the first 21 months of the pandemic to address the following questions: (1) what was the

acceptance rate of these submissions? and (2) after publication, how often did the preprint server note the published article?

**Design** Data on all submissions from March 1, 2020, through January 21, 2022, that appeared on preprint servers were collected. Editorial decisions (reject without external review, reject after external review or statistical review, and accepted for publication) were reported. Finally, for those submissions that the journal did not publish but for which it did perform a statistical review, a Google Scholar search was performed to determine subsequent publication in another indexed journal.

**Results** Of all manuscripts submitted between March 1, 2020, and January 21, 2022, a total of 362 manuscripts were posted in a preprint archive prior to submission to *Annals of Internal Medicine*. Of these, 337 manuscripts (93.6%) were rejected, 23 manuscripts (6.3%) were published, and 2 manuscripts were pending final decision. Additionally, 247 manuscripts (68.2%) were rejected without external review and 115 manuscripts (31.8%) were sent for peer review. Of those sent for peer review, 34 manuscripts (29.5%) progressed to statistical review. Of those sent for statistical review, 11 manuscripts (32.4%) were rejected after statistical review. A Google Scholar search showed that 2 manuscripts rejected after statistical review were published in other journals. Among manuscripts published to preprint servers, a link to the peer-reviewed, published version was provided on the server for 13 manuscripts (56.5%). Acceptance and rejection rates were comparable to submissions that did not appear on a preprint server.

**Conclusions** There are theoretical advantages to posting non-peer-reviewed preprints of clinical research, particularly during public health crises, such as a pandemic. However, subsequent vetting and publication in peer-reviewed journals can help to avoid dissemination of misinformation. Unfortunately, a large proportion of submissions posted as preprints were not found suitable for publication in a clinically influential, peer-reviewed journal. When published, the preprint archive did not always acknowledge the subsequently published article.

## Reference

1. Cabanac G, Oikonomidi T, Boutron I. Day-to-day discovery of preprint-publication links. *Scientometrics*. 2021;126(6):5285–5304. doi:10.1007/s11192-021-03900-7

<sup>1</sup>*Annals of Internal Medicine*, American College of Physicians, Philadelphia, PA, USA, jjackson@acponline.org

**Conflict of Interest Disclosures** Christine Laine is a member of the Peer Review Congress advisory board but was not involved in the review or decision for this abstract. No other disclosures reported.

---

## Assessment of the Pros and Cons of Posting Preprints Online Before Submission to a Double-Anonymous Review Process in Computer Sciences

Charvi Rastogi,<sup>1</sup> Ivan Stelmakh,<sup>1</sup> Xinwei Shen,<sup>2</sup> Marina Meila,<sup>3</sup> Federico Echenique,<sup>4</sup> Shuchi Chawla,<sup>5</sup> Nihar B. Shah<sup>1</sup>

**Objective** Authors posting preprints online before review in double-anonymous peer review is a widely debated issue of policy as well as authors' personal choice.<sup>1,2</sup> Authors in a disadvantaged group posting preprints online can gain visibility but may lose the benefits of double-anonymous review.<sup>1</sup> In this work, this debate was substantiated by quantifying (1) how frequently reviewers deliberately search for their assigned paper online and (2) the correlation between the visibility of preprints posted online and the prestige of the associated authors' affiliations.

**Design** Surveys were conducted in 2 top-tier computer science conferences that review full papers, are terminal publication venues, and are considered on par with journals: the 2021 Association for Computing Machinery's Conference on Economics and Computation (EC) and 2021 International Conference on Machine Learning (ICML). To address this work's first objective, after the initial review period, an anonymized survey was conducted (by invitation and with validation to protect from spurious responses), with the following question sent to each reviewer: "During the review process, did you search for any of your assigned papers on the internet?" It was clarified that accidental discovery of a paper on the internet (eg, through searching for related works) did not count as a positive case for this question. To address the second objective, a list was compiled of papers that were submitted to ICML or EC and found available online before review. Relevant reviewers were surveyed during the review process on whether they had seen some of these papers online outside of a reviewing context. The visibility of a paper was set to 1 if the reviewer responded yes; otherwise, it was set to 0. To quantify prestige, each paper was assigned a prestige metric based on the ranking of the authors' affiliations in widely used world rankings such as QS (Quacquarelli Symonds) rankings and Computer Science rankings. Finally, a Kendall τ b correlation coefficient was computed between papers' visibility and prestige metrics.

**Results** For the first objective, more than 35% of the survey respondents self-reported searching online for their assigned paper in ICML and EC. A weakly positive correlation was observed in the visibility of preprints posted online and their prestige metric, which was statistically significant in ICML but not in EC. To interpret the correlation, the mean visibility of papers with high prestige was compared with that of the remaining papers (**Table 73**). The results for the second objective were based on preprints posted online. Further analysis was conducted to account for this fact, wherein correlation between papers' prestige metric and final decision was computed for papers that were and were not posted online.

**Table 73. Results of the Experiment**

Conference	EC 2021	ICML 2021
<b>Survey 1</b>		
Response rate, % (No./total No.)	51.1 (97/190)	16.0 (753/4699)
Fraction of reviewers deliberately searching for their assigned paper online, % (No./total No.)	42.2 (41/97)	35.7 (269/753)
<b>Survey 2</b>		
Response rate, % (No./total No.)	55.8 (449/805)	100 (7594/7594)
Mean visibility of submissions with online preprints, % (No./total No.)	20.5 (92/449)	8.4 (635/7594)
Correlation between prestige and visibility of paper <sup>a</sup>	0.05	0.06
P value for correlation	.10	<.001
Mean visibility of submissions with online preprints with high prestige metric (rank <10), % (No./total No.)	22.8 (59/259)	10.9 (247/2266)
Mean visibility of submissions with online preprints with low prestige metric (rank ≥10), % (No./total No.)	17.4 (33/190)	7.3 (388/5328)

Abbreviations: EC, Association for Computing Machinery's Conference on Economics and Computation; ICML, International Conference on Machine Learning.

<sup>a</sup>A positive correlation in survey 2 indicates that the visibility of a paper increases as the prestige of the associated affiliation increases.

**Conclusions** Based on this work,<sup>3</sup> peer review organizers and authors posting preprints online should account for the finding that a substantial fraction of reviewers searched for their assigned papers online in EC 2021 and ICML 2021.

## References

1. Tomkins A, Zhang M, Heavlin WD. Reviewer bias in single- versus double-blind peer review. *Proc Natl Acad Sci USA*. 2017;114(48):12708-12713. doi:10.1073/pnas.1707323114
2. Blank RM. The effects of double-blind versus single-blind reviewing: experimental evidence from the American Economic Review. *Am Econ Rev*. 1991;81(5):1041-1067.
3. Rastogi C, Stelmakh I, Shen X, et al. To ArXiv or not to ArXiv: a study quantifying pros and cons of posting preprints online. *arXiv*. Preprint posted online March 31, 2022. Accessed July 13, 2022. <https://arxiv.org/abs/2203.17259>

<sup>1</sup>Carnegie Mellon University, Pittsburgh, PA, USA; <sup>2</sup>edu; <sup>3</sup>Hong Kong University of Science and Technology, Hong Kong;

<sup>3</sup>University of Washington, Seattle, WA, USA; <sup>4</sup>California Institute of Technology, Pasadena, CA, USA; <sup>5</sup>The University of Texas at Austin, Austin, TX, USA

**Conflict of Interest Disclosures** Ivan Stelmakh reports a collaboration with Google Research through a summer research internship. Charvi Rastogi reports a collaboration with IBM Research New York through a summer research internship. These interests did not play any role in the submitted research.

**Funding/Support** This work was supported by the US National Science Foundation (NSF) in part by NSF CAREER award 1942124, which supports research on the fundamentals of learning from people with applications to peer review. Marina Meila was supported by NSF MMS Award 2019901. Federico Echenique was supported by NSF awards SES 1558757 and CNS 1518941.

**Role of the Funder/Sponsor** The funder did not play a role in any of the following: design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Acknowledgment** The authors gratefully acknowledge Tong Zhang, program co-chair of the International Conference on Machine Learning 2021 jointly with Marina Meila, for his contribution in designing the workflow, designing the reviewer questions, facilitating access to relevant summaries of anonymized data, and supporting the polling of the reviewers, as well as for many other helpful discussions and interactions; and Hanyu Zhang, workflow co-chair jointly with Xinwei Shen, for his contributions to the workflow and many helpful interactions.

**Additional Information** Nihar B. Shah is a co-corresponding author.

## Quality of Reporting

### Reporting of Retrospective Registration in Clinical Trial Publications

Martin Haslberger,<sup>1</sup> Stefanie Gestrich,<sup>1</sup> Daniel Strech<sup>1</sup>

**Objective** Preregistration of clinical research has been widely implemented and advocated for many reasons: to detect and mitigate publication bias, selective reporting, and undisclosed changes in determination of primary and secondary outcomes. Prospective registration allows for public scrutiny of trials to identify research gaps and to support the coordination of efforts by preventing unnecessary duplication. Retrospective registration undermines many of these reasons but is commonly found. This study provided a comprehensive analysis of retrospective registration and the reporting thereof in publications, as well as associated factors, based on a validated data set of trial registrations and results publications from Germany from 2009 to 2017.

**Design** The study used a validated and previously published data set<sup>1,2</sup> of trials registered on ClinicalTrials.gov or DRKS (German Clinical Trials Register), led by German University Medical Centers, completed between 2009 and 2017, and with at least 1 peer-reviewed results publication. From all results publications of retrospectively registered trials, all registration statements, including mentions and justifications of retrospective registration, were extracted. Associations between key trial variables and different registration and reporting practices were analyzed.

**Results** Based on an analysis of 1030 retrospectively registered clinical trials, 2.0% (21) explicitly reported the retrospective registration in the abstract and 3.3% (34) in the full text. In 2.3% (24) of publications, a justification or explanation was provided in the full text. Analyses are ongoing; full results will be presented at the conference, including a qualitative analysis of the reasons given for retrospective registration, as well as trends over time and exploratory analyses of the associations between retrospective registration and other reporting practices, such as registration

number reporting and cross-registration practices between different registries.

**Conclusions** Disclosure of retrospective registration would be a positive signal for rigor, as the registrants would feel it critical to transparently report this limitation. However, only a small number of retrospectively registered studies reported the retrospective nature of the registration. Lack of disclosure might lead readers to wrongly interpret the registration as a quality criterion that, in the case of a retrospective registration, rather describes a concern. This study provided a detailed analysis of this issue.

## References

1. Riedel N, Wieschowski S, Bruckner T, et al. Results dissemination from completed clinical trials conducted at German university medical centers remained delayed and incomplete: the 2014-2017 cohort. *J Clin Epidemiol*. 2022;144:1-7. doi:10.1016/j.jclinepi.2021.12.012
2. Wieschowski S, Riedel N, Wollmann K, et al. Result dissemination from clinical trials conducted at German university medical centers was delayed and incomplete. *J Clin Epidemiol*. 2019;115:37-45. doi:10.1016/j.jclinepi.2019.06.002

<sup>1</sup>Berlin Institute of Health at Charité, Universitätsmedizin Berlin, QUEST Center for Responsible Research, Berlin, Germany, martin.haslberger@bih-charite.de

**Conflict of Interest Disclosures** Daniel Strehl is a member of the Sanofi Advisory Bioethics Committee and receives an honorarium for his contribution to meetings. No other disclosures were reported.

**Funding/Support** The project was funded from QUEST departmental resources.

**Additional Information** Daniel Strehl is a co-corresponding author.

---

## Reporting of Methods Used to Ascertain Adverse Events of Special Interest (AESI) and Adverse Events Newly Signaled After Marketing Authorization of Drugs Approved Between 2018 and 2019

Kyungwan Hong,<sup>1</sup> Anisa Rowhani-Farid,<sup>1</sup> Francis B. Palumbo,<sup>1</sup> John H. Powers III,<sup>1,2</sup> Linda Wastila,<sup>1</sup> Peter Doshi<sup>1</sup>

**Objective** Randomized clinical trials (RCTs) ascertain adverse events using 2 main approaches: (1) the nonsystematic approach collects data passively asking nontargeted questions relying on participants' spontaneous responses, and (2) the systematic approach proactively collects data using standardized solicitation tools (eg, questionnaires).<sup>1</sup> The ascertainment approach may affect the recorded rate of adverse events, affecting the interpretation of safety results from trials. Using publicly available trial documents, reported adverse event ascertainment methodologies used by regulatory agencies for marketing approval of new drugs were descriptively assessed.

**Design** All new molecular entities and biologic license applications approved by the US Food and Drug Administration (FDA) in 2018 to 2019 were screened across a variety of indications. Where publicly available, clinical study reports, study protocols, and case report forms were obtained from trial registries, journals, and regulatory agencies' websites. Reported ascertainment methods of 2 categories of adverse events were screened: adverse event of special interest (AESI) (adverse event types often predefined by investigators/sponsors to facilitate systematic assessment) and adverse events signaled by the European Medicines Agency Pharmacovigilance Risk Assessment Committee (EMA PRAC) within 12 months of market authorization (adverse-event types that conceivably could have been captured in preapproval trials). One researcher extracted the data. A second researcher verified a 10% random sample. Discrepancies were resolved through discussion.

**Results** A total of 107 new molecular entities and biologic license applications approved in 2018 to 2019 were screened, and 31 drugs (29.0%) had publicly available underlying trial documents. Of 64 pivotal trials, AESIs were described in 54 clinical study reports (84.4%; 95% CI, 73%-92%), 33 study protocols (51.6%; 95% CI, 39%-64%), and 10 journal publications (15.6%; 95% CI, 8%-27%). A total of 322 AESIs were identified from trial documents. The median number of AESIs per trial was 4. Mainly using diagnostic measurement tools and laboratory assessments, 71.4% of AESIs (230 of 322; 95% CI, 66%-76%) were systematically ascertained, and 9.3% (30 of 322; 95% CI, 6%-13%) were nonsystematically ascertained. The ascertainment method of 19.3% (62 of 322; 95% CI, 15%-24%) was unclear (**Table 74**).<sup>2</sup> In 38 of 54 trials (70.4%; 95% CI, 56%-82%), at least 1 AESI was nonsystematically ascertained and/or had unclear ascertainment methodology. From EMA PRAC reports, 1 newly signaled adverse event was identified and systematically ascertained using targeted questions in case report forms.

**Conclusions** The ascertainment methodology for approximately 20% of AESIs was unable to be identified, even with access to underlying trial documents. Most systematically ascertained AESIs were laboratory abnormalities rather than direct measures of patient health. The lack of adequate reporting impedes the accurate interpretation of a drug's adverse event profile. Given that regulatory agencies expect more rigorous data collection for AESIs,<sup>3</sup> the finding that approximately 10% of AESIs were assessed nonsystematically suggests a need for more rigorous methodology to capture adverse events and increased regulatory oversight. Further investigation is needed to explore reporting of ascertainment methodologies used for other types of adverse events.

## References

1. Mayo-Wilson E, Fusco N, Li T, et al. Harms are assessed inconsistently and reported inadequately, part 1: systematic adverse events. *J Clin Epidemiol*. 2019;113:20-27. doi:10.1016/j.jclinepi.2019.04.022

**Table 74. Characteristics and Data Collection Methods of 322 Examined AESIs**

Characteristic	No. (%)
No. of AESIs per trial, median (IQR)	4 (1.75-8.00)
<b>General symptoms (eg, dry mouth, itchiness, cough)</b>	
Yes	65 (20.2)
<b>Adverse events classifications<sup>a</sup></b>	
Blood and lymphatic system	37 (11.5)
Cardiovascular	36 (11.2)
Gastrointestinal	19 (5.9)
General and other miscellaneous disorders	25 (7.8)
Hepatobiliary	17 (5.3)
Immune system	12 (3.7)
Infections	35 (10.9)
Laboratory abnormalities	35 (10.9)
Musculoskeletal	10 (3.1)
Neoplasms (benign, malignant, or unspecific)	25 (7.8)
Nervous system and psychiatric disorders	23 (7.1)
Ophthalmic	15 (4.7)
Kidney and endocrine disorders	17 (5.3)
Reproductive system and other congenital disorders	3 (0.9)
Respiratory	3 (0.9)
Skin related	10 (3.1)
<b>Ascertainment</b>	
Systematically ascertained	230 (71.4)
Nonsystematically ascertained	30 (9.3)
Unclear	62 (19.3)
<b>Data collection methodology</b>	
Diagnostic measurement tools (eg, questionnaires)	97 (30.1)
Laboratory assessments	101 (31.4)
Additional data collection using CRFs	32 (9.9)
Not specifically described other than generic statement	92 (28.6)
<b>Methods used to code and grade the adverse events</b>	
MedDRA alone	85 (26.4)
MedDRA and CTCAE	156 (48.4)
MedDRA and RCTC	81 (25.2)

Abbreviations: AESI, adverse event of special interest; CRF, case report form; CTCAE, Common Terminology Criteria for Adverse Events; MedDRA, Medical Dictionary for Regulatory Activities; RCTC, Rheumatology Common Toxicity Criteria.

<sup>a</sup>The classification categories were modified from the MedDRA's system organ classes. The information on each adverse event's MedDRA classification was from the National Center for Biomedical Ontology's BioPortal.<sup>2</sup>

2. Whetzel PL, Noy NF, Shah NH, et al. BioPortal: enhanced functionality via new web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.* 2011;39:W541-W545. doi:10.1093/nar/gkr469

3. US Food and Drug Administration. E19: optimisation of safety data collection. Accessed January 13, 2022. <https://www.fda.gov/media/128313/download>

<sup>1</sup>Department of Pharmaceutical Health Services Research, University of Maryland School of Pharmacy, Baltimore, MD, USA, hong.kyungwan@umaryland.edu; <sup>2</sup>Department of Medicine, George Washington University School of Medicine, Washington, DC, USA

**Conflict of Interest Disclosures** Kyungwan Hong reported receiving a financial assistance award from the US Food and Drug Administration (FDA) of the US Department of Health and Human Services (HHS) outside the submitted work. Kyungwan Hong, Anisa Rowhani-Farid, and Peter Doshi reported receiving salaries from the RIAT Support Center, funded by the Laura and John Arnold Foundation. John H. Powers III reported receiving consulting fees from Arrevus, Eicos, Eli Lilly, Evofem, Eyecheck, Fuji, Gilead, GlaxoSmithKline, Johnson & Johnson, Microbion, OPKO, Otsuka, Resolve, Romark, Shinogi, SpineBioPharma, Utility, and Vir outside the submitted work. Peter Doshi reported receiving travel funds from the European Respiratory Society and Uppsala Monitoring Center; grants from the FDA (through the University of Maryland and the Center of Excellence in Regulatory Science and Innovation), the Laura and John Arnold Foundation, the American Association of Colleges of Pharmacy, the Patient-Centered Outcomes Research Institute, the Cochrane Methods Innovations Fund, and the UK National Institute for Health Research; reported being an unpaid Innovation in Medical Evidence Development and Surveillance steering committee member at the Reagan-Udall Foundation for the FDA; and being an editor at *The BMJ*. No other disclosures were reported.

**Additional Information** Peter Doshi is a co-corresponding author.

## Quality of Trials

### Geographical Scope of Randomized Clinical Trials From Africa

Folafoluwa Olutobi Odetola,<sup>1</sup> Marisa L. Conte<sup>2</sup>

**Objective** Africa ranks second of the 7 continents in population and geographical area. Reports of research from Africa are often labeled as being African in scope, a dubious claim given the low likelihood of analyzing data representative of 1.4 billion people. Accurate reporting of the scope of research studies in Africa avoids mischaracterization and bias that could have significant implications on health care resource distribution, research funding, and policy making. This study characterized the geographical scope of reports of randomized clinical trials (RCTs) conducted in Africa, because RCTs are most likely to inform clinical practice and policy.

**Design** A tailored PubMed search was developed using Medical Subject Headings and keywords to represent *Africa*, *African*, and *RCTs*. Citations from January 1, 1968, to February 25, 2022, were included. Titles and abstracts were reviewed against established inclusion and exclusion criteria followed by a full-text review of included citations and data extraction. Based on the number of countries, studies were categorized as national (1 country), binational (2 countries), trinational (3 countries), multinational ( $\geq 4$  countries), and continentwide (multinational, inclusive of all 5 subregions: North, Southern, Central, East, and West).

**Results** A total of 285 RCTs met inclusion criteria, and only 1 was continentwide. Of the 54 African countries, 14 (25%) were not represented in the published RCTs. The top 3 countries were in East Africa, namely, Kenya (80), Uganda (79), and Tanzania (56). Studies were reported as African (201 [71%]), sub-Saharan African (45 [16%]), West African (16 [6%]), East African (14 [5%]), Southern African (7 [2%]), and East + Central or West + Central African (2 [<1%]). The median number of countries included by reported scope was 2 (IQR, 1-3) overall, and similar in African and sub-Saharan African studies; 3.5 (IQR, 2-4) in East African studies; 1 (IQR, 1-2) in West African studies; and 2 (IQR, 2-2) in Southern African studies. A total of 124 studies (44%) were national, 66 (23%) were binational, 33 (11%) were trinational, and 62 (22%) were multinational. Of 201 studies reported as African, 91 (45%) were national, 49 (25%) were binational, 21 (10%) were trinational, 39 (19%) were multinational, and 1 (<1%) was continentwide. Among the 45 (16%) sub-Saharan African studies, 20 (45%) were national, 6 (13%) were binational, 9 (20%) were trinational, and 10 (22%) were multinational. Among the 14 (5%) East African studies, 2 (14%) were national, 3 (22%) were binational, 2 (14%) were trinational, and 7 (50%) were multinational. Of the 16 studies (6%) conducted in West Africa, 10 (62%) were national, 3 (19%) were binational, and 3 (19%) were multinational. Among the 7 RCTs from Southern Africa, 1 (14%) was national, 5 (72%) were binational, and 1 (14%) was trinational.

**Conclusions** The scope of RCTs from Africa is rarely continentwide, and 1 in 4 countries is not included.

<sup>1</sup>Department of Pediatrics, University of Michigan, Ann Arbor, MI, USA, fodenola@med.umich.edu; <sup>2</sup>Taubman Health Sciences Library, University of Michigan, Ann Arbor, MI, USA

**Conflict of Interest Disclosures** None reported.

## Reporting Guidelines

### Development of the Standards for Reporting Subtyping Studies (StaRSS) Reporting Guideline

Seyed-Mohammad Fereshtehnejad,<sup>1,2</sup> Connie Marras,<sup>3,4</sup> David Moher,<sup>5,6</sup> Tiago Mestre,<sup>1</sup> for the International Parkinson and Movement Disorder Society Task Force on Parkinson's Disease's Subtypes

**Objective** In the evolving era of precision medicine, a subtyping study attempts to define subgroups of a disease entity wherein certain individuals with the disease share clinical characteristics (eg, biomarker profile, clinical manifestations, and prognosis) distinct from others with the disease. There is an increase in the number of studies on disease subtyping in the past 40 years, with a rapid surge in the last decade, yet there is no standardized reporting guideline or appraisal checklist for subtyping studies to promote reporting quality. This new guideline aimed to improve the accuracy and completeness of reporting of studies of disease subtyping and to aid readers and peer

reviewers in assessing the potential for bias in a subtyping study.

**Design** Members of the International Parkinson and Movement Disorder Society Task Force on Parkinson's Disease's Subtypes, having performed a systematic review of subtyping studies of Parkinson disease and found significant methodologic shortcomings of published subtyping research,<sup>1</sup> followed up with a systematic search of the Enhancing the Quality and Transparency of Health Research (EQUATOR) Network database and others to find published guidelines about subtyping research, with no results.<sup>2</sup> In response, the Task Force convened expert panel meetings to develop a Standards for Reporting Subtyping Studies (StaRSS) reporting guideline,<sup>1</sup> following the steps introduced in the guidance for developers of health research reporting guidelines.<sup>3</sup> To do so, members of the expert panel created new items relevant to the context of subtyping using Parkinson disease as a model. The items were approved after multiple revisions in expert panel meetings.

**Results** The reporting guideline comprises 37 items, including numerous novel topics unique to the concepts and various aspects of a subtyping study mainly related to the Methods and Results sections of subtyping studies

(**Table 75**). For the other sections, the Task Force adopted relevant items from existing general guidelines, namely, the Standards for Reporting of Diagnostic Accuracy (STARD) and the Reporting of Observational Studies in Epidemiology (STROBE), with minor adjustments.

**Conclusions** StaRSS is a consensus guideline that describes essential elements of the design, outcome assessment, execution of statistical tests, and reporting of subtype research findings and could guide authors to properly conduct a subtyping study. Adoption of the StaRSS checklist by biomedical journals might improve the quality of reporting and appraisal of subtyping studies by peer reviewers. To achieve these goals, future studies are needed to assess the utility of the StaRSS checklist through panel discussions of the experts in subtyping studies of various fields of biomedicine.

### References

1. Mestre TA, Fereshtehnejad SM, Berg D, et al. Parkinson's disease subtypes: critical appraisal and recommendations. *J Parkinsons Dis.* 2021;11(2):395-404. doi:10.3233/JPD-202472
2. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Lancet.* 2008;371(9619):1149-1150. doi:10.1016/S0140-6736(08)60505-X
3. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med.* 2010;7(2):e1000217. doi:10.1371/journal.pmed.1000217

<sup>1</sup>Division of Neurology, Department of Medicine, University of Ottawa, Ottawa, ON, Canada, sm.fereshtehnejad@ki.se; <sup>2</sup>Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences

**Table 75. Standards for Reporting Subtyping Studies (StaRSS)<sup>a</sup>**

Section/topic	No.	Item
<b>Title</b>		
	1	Identification as a subtyping study
<b>Abstract</b>		
	2	Structured summary of study design, methods, results, and conclusions (reporting the most important measure of between-group comparison)
<b>Introduction</b>		
Background/ rationale	3	Scientific and clinical background, including the intended use and clinical role of the index subtyping method
Objectives	4	Study objectives and hypotheses (includes subtyping as the main objective)
	5	Subtyping approach/purpose (eg, prognostication and pathophysiology)
<b>Methods</b>		
Study design	6	Whether data collection was planned for the subtyping study (prospective) or subtyping was performed on data collected for another purpose (retrospective)
Participants	7	Eligibility criteria
	8	Basis on which eligible participants were identified (eg, symptoms, definition criteria, and inclusion in registry)
	9	Disease stage at baseline (eg, de novo, drug-naïve, and advanced)
	10	Where and when potentially eligible participants were identified (ie, setting, location, recruitment source, and dates)
Sampling method	11	Whether participants formed a consecutive, random, or convenience series
	12	Intended sample size and how sample size was determined
Subtyping method	13	Subtyping approach: whether study applies an a priori or hypothesis-free (data-driven) approach
	14	Subtyping tools and variables: whether study uses a clinical phenotyping, biomarker subtyping (eg, pathological, molecular, genetic, imaging, and serum markers), or integrative method
Outcome assessment	15	Clear definition of the outcome variable(s) or external variables
	16	Longitudinal follow-up to compare prognoses between subtypes (ie, duration, setting, and methods of data collection)
	17	Whether subtype population data were available to the clinicians or researchers assessing the outcome in longitudinal studies
Statistical methods	18	Statistical methods for creating the subtypes (eg, clustering, machine-learning, and latent class analysis)
	19	How missing data were handled

and Society (NVS), Karolinska Institutet, Stockholm, Sweden; <sup>3</sup>Movement Disorders Clinic, Toronto Western Hospital, University Health Network, Toronto, ON, Canada; <sup>4</sup>Division of Neurology, Department of Medicine, University of Toronto, Toronto, ON, Canada; <sup>5</sup>Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada;

<sup>6</sup>Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

**Conflict of Interest Disclosures** Seyed-Mohammad Fereshtehnejad received honoraria from the European Science

Section/topic	No.	Item
	20	If applicable, how loss to follow-up was addressed
	21	Post hoc analyses of variability between subtypes (eg, univariate analysis and multivariable analysis)
	22	How type I error inflation due to multiple comparisons was handled
<b>Results</b>		
Descriptions	23	Flow of participants from baseline to follow-up visits using a diagram (ie, completeness of follow-up)
	24	Baseline demographic and clinical characteristics of participants in each subtype
Subtype distinction	25	Differences in the main subtyping features used to create or define subtypes
	26	Distinguishing external features that were not included in the definition or exploration of subtypes
	27	Comparing longitudinal trajectories of the outcome variable(s) between subtypes
Subtyping stability	28	Analyzing and reporting longitudinal stability of subtypes during follow-up (eg, whether the subtype population changes over time (subtype shifting))
Validation	29	Cross-validation (training and testing sets); reproducibility in an external cohort
<b>Discussion</b>		
Key results	30	Summary of key results with reference to study objectives
Interpretations	31	Cautious overall interpretation of results considering objectives, limitations, reproducibility of analyses, comparisons with other subtyping methods, and other relevant evidence
Limitations	32	Study limitations, including sources of potential bias (eg, selection bias and information bias) and statistical uncertainty
Generalizability	33	Whether the subtyping method is generalizable to all individuals with the reference disease or condition
Applicability	34	Clinical applicability of the subtyping method (eg, cost, time requirement, accessibility, and invasiveness)
<b>Other information</b>		
	35	Clinically applicable algorithm or guideline for subtyping in real-world clinical practice and research settings, either in the main article or in the supplementary materials (knowledge translation)
	36	Where the full study protocol can be accessed
	37	Sources of funding and other support; role of funders

<sup>a</sup>Adapted from the Standards for Reporting Diagnostic Accuracy Studies (STARD) checklist and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist for observational cohort studies.

Foundation. Connie Marras is a consultant for Grey Matter Technologies and receives financial compensation as a steering committee member from the Michael J. Fox Foundation and the Parkinson's Foundation (USA); she received research grants from the Michael J. Fox Foundation, Canadian Institutes of Health Research, International Parkinson and Movement Disorder Society, and the Weston Brain Institute and as a site investigator for clinical studies sponsored by Theravance Inc, the Michael J. Fox Foundation, the Parkinson's Foundation, and Centogene. David Moher is an associate director of the Peer Review Congress Advisory Board but was not involved in the review or decision for

this abstract. Tiago Mestre reports speaker honoraria from AbbVie and the International Parkinson and Movement Disorder Society; consultancies from CHDI Foundation/management and advisory board from AbbVie; Biogen, Roche, nQ Medical, Medtronic, Sunovion, and Valeo Pharma; and research funding from EU Joint Programme—Neurodegenerative Disease Research, University of Ottawa Brain and Mind Research Institute, Ontario Research Fund, Canadian Institutes of Health Research, the Michael J. Fox Foundation, Parkinson Canada, Parkinson’s Disease Foundation/Parkinson Study Group, LesLois Foundation, PSI Foundation, and the Parkinson Research Consortium.

**Acknowledgments** We would like to thank all the members of the Parkinson’s disease subtyping task force at the International Movement Disorder Society.

## Reproducible Research

### Assessment of Minimum False-Positive Risk of Primary Outcomes After Reducing the Nominal P Value Threshold for Statistical Significance From .05 to .005 in Anesthesiology Randomized Clinical Trials

Philip M. Jones,<sup>1,2,3</sup> Zachary Chuang,<sup>1</sup> Janet Martin<sup>1,2,3</sup> Derek Nguyen,<sup>1</sup> Jordan Shapiro,<sup>1</sup> Penelope Neocleous<sup>1</sup>

**Objective** A primary reason for reproducibility concerns in the biomedical literature may be that many published articles reporting statistically significant findings do not represent real effects.<sup>1,2</sup> Several solutions have been postulated to mitigate the risks associated with false-positive findings.<sup>1,2</sup> This study sought to determine the ramifications of lowering the nominal *P* value for statistical significance from .05 to .005 and assessed the minimum false-positive risk (minFPR) for primary outcomes in anesthesiology randomized clinical trials (RCTs). These proposals have been explored in other fields, but the metrics have not been quantified for anesthesiology.

**Design** This cross-sectional descriptive study aimed to determine these metrics for RCTs published in the top general anesthesiology journals, defined by impact factor. The target journals were *Anesthesia*, *Anesthesia & Analgesia*, *Anesthesiology*, *British Journal of Anaesthesia*, *Canadian Journal of Anesthesia*, *European Journal of Anaesthesiology*, and *Journal of Clinical Anesthesia*. The Cochrane Highly Sensitive Search Strategy was used to identify RCTs in MEDLINE. All superiority RCTs published between January 1, 2019, and March 15, 2021, comparing 2 groups with at least 1 primary outcome were included. Study screening and data extraction were performed in duplicate. *P* values for primary outcomes were extracted and the percentage of RCTs that would maintain statistical significance at a threshold of  $P < .005$  was determined. For these outcomes, minFPRs were calculated assuming 1:1 prior odds of an intervention being effective, using previously recommended methods.<sup>3</sup> Study-level characteristics predicting maintenance of statistical significance at  $P < .005$  and minFPRs were computed using logistic and median regression, respectively.

**Results** After searching, deduplication, and screening, 318 RCTs were included. The median (IQR) sample size was 80 (52-130) and did not differ significantly across journals. The majority of RCTs (273 of 318 [86%]) were single-center studies. *P* values below .05 occurred in 205 of 318 RCTs (64%) (by journal, this ranged from 44% to 77%). Of these 205, 119 (58%; 95% CI, 51%-65%) maintained statistical significance at the  $P < .005$  threshold. The mean (SD) minFPR was 22% (20%) (by journal, this ranged from 16% to 33%). Violin plots for *P* values and minFPRs by journal are shown in **Figure 27**. With minFPR<sub>50</sub> (ie, minFPR assuming a prior probability of 50%) constrained to RCTs with  $P < .005$ , the mean (SD) was 2% (1.2%).

**Conclusions** Approximately 42% of primary outcomes in anesthesiology RCTs would lose statistical significance under a more stringent *P* value threshold of .005. These primary outcomes carry a minimum false-positive risk of 22%. The adoption of the  $P = .005$  threshold for statistical significance could reduce the minFPR to just 2%. These results call a large portion of anesthesiology RCTs into question and provide impetus to improve study design, analysis, and reporting methods to reduce false-positives and improve reproducibility.

## References

1. Niven DJ, McCormick TJ, Straus SE, et al. Reproducibility of clinical research in critical care: a scoping review. *BMC Med*. 2018;16:26. doi:10.1186/s12916-018-1018-6
2. Colquhoun D. The false positive risk: a proposal concerning what to do about *P*-values. *Am Stat*. 2019;73:192-201. doi:10.1080/00031305.2018.1529622
3. Sellke T, Bayarri MJ, Berger JO. Calibration of *P* values for testing precise null hypotheses. *Am Stat*. 2001;55:62-71. doi:10.1198/000313001300339950

<sup>1</sup>Schulich School of Medicine & Dentistry, University of Western Ontario, London, ON, Canada, philip.jones@lhsc.on.ca;

<sup>2</sup>Department of Anesthesia & Perioperative Medicine, University of Western Ontario, London, ON, Canada; <sup>3</sup>Department of Epidemiology & Biostatistics, University of Western Ontario, London, ON, Canada

**Conflict of Interest Disclosures** Philip M. Jones is deputy editor in chief at the *Canadian Journal of Anesthesia*. No other disclosures were reported.

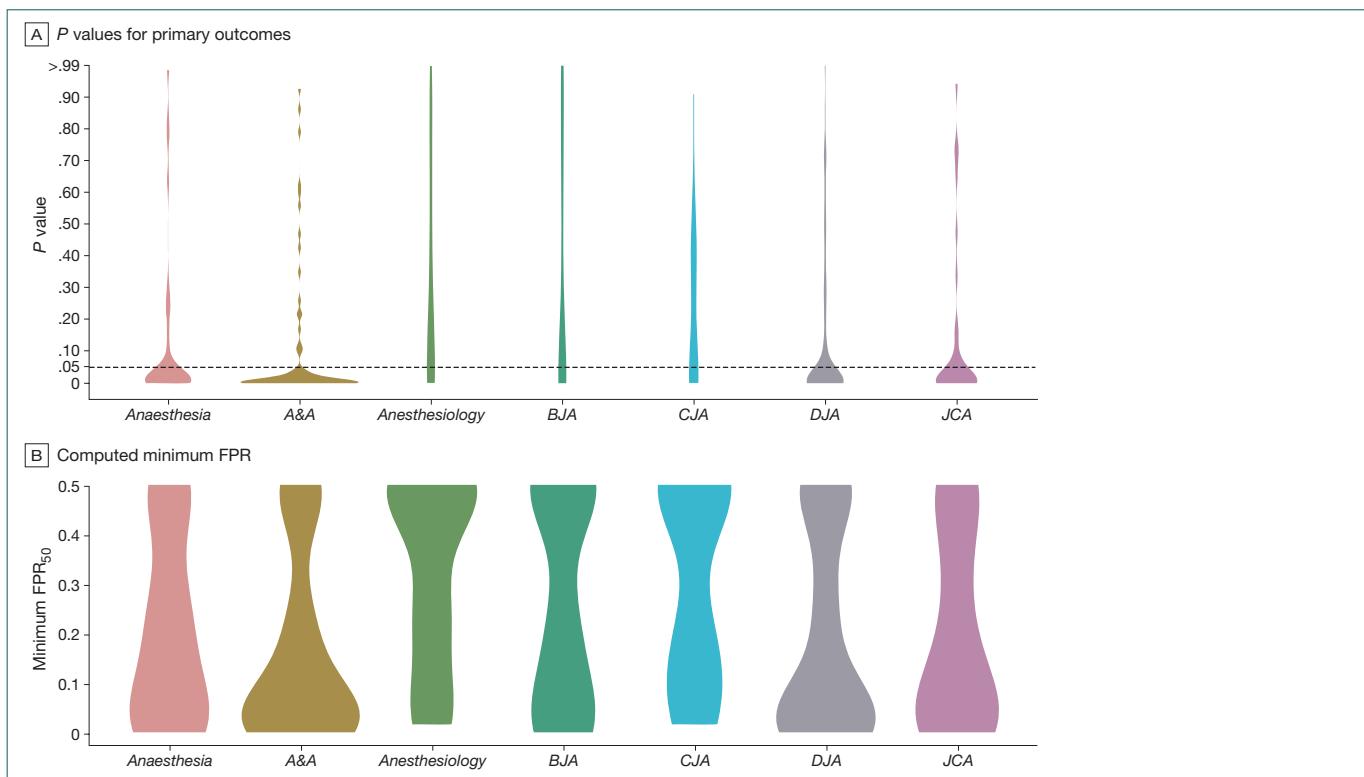
**Funding/Support** Research time for Philip M. Jones and Janet Martin was provided by the Department of Anesthesia & Perioperative Medicine at the University of Western Ontario, London, ON, Canada.

**Role of the Funder/Sponsor** The Department of Anesthesia & Perioperative Medicine was not involved in the design or conduct of the study, nor the preparation, review, approval or submission of the abstract for presentation.

**Additional Information** This study was registered on March 15, 2021, with Open Science Framework (doi:10.17605/OSF.IO/H8KBZ).

**Acknowledgments** We are grateful to David Colquhoun, who reviewed a presubmission draft of the manuscript.

**Figure 27. P Value and Minimum False-Positive Risk (FPR) Distribution for Randomized Clinical Trials (RCTs) Published in Major Anesthesiology Journals**



A, Violin plot showing the smoothed distribution (probability density) of  $P$  values for primary outcomes published in 7 major anesthesiology journals from January 1, 2019, to March 15, 2021. The width of the plot at any particular  $P$  value is proportional to the frequency of that  $P$  value for that journal. The horizontal dashed line shows  $P = .05$ . B, Violin plot for the computed minimum false-positive risk assuming a prior probability of 50% ( $FPR_{50}$ ) for the same RCTs. A&A indicates *Anesthesia & Analgesia*; BJA, *British Journal of Anaesthesia*; CJA, *Canadian Journal of Anesthesia*; EJA, *European Journal of Anaesthesiology*; and JCA, *Journal of Clinical Anesthesia*.

## Retractions

### Analysis of Articles Retracted Because of Conflicts of Interest in the Retraction Watch Database

Ružica Bočina,<sup>1,2</sup> Antonija Mijatović,<sup>1,2</sup> Ana Marušić<sup>1,2</sup>

**Objective** The Retraction Watch Database reported in 2018 that conflict of interest (COI) was not a common sole reason for a retraction.<sup>1</sup> This study assesses the characteristics of articles in the Retraction Watch Database that have COI as a reason for retraction.

**Design** All retractions from Retraction Watch Database that had COI as a reason for retraction were analyzed. Using Scopus Application Programming Interface, data on the types of articles retracted, fields of retracted articles, number of authors and institutions, time between the publication of the article and the retraction, first corresponding author's country, and other reasons for retraction were collected. The Retraction Watch Database was also manually searched to identify articles with expression of concern, errata, or corrigendum notices for COI. The linguistic content of notices was compared using the Linguistic Inquiry and Word Count program, and t-distributed stochastic neighbor embedding technique (t-SNE) was used to visualize high-dimensional data by placing each data point in a 2-dimensional map.<sup>2</sup>

**Results** Of 194 articles tagged with COI as a reason for retraction from 1991 to 2021, COI was the only reason for retraction for 26 (13%). The most common additional reasons for retraction were fake peer review ( $n = 34$ ), investigation by journal/publisher ( $n = 29$ ), concerns/issues about authorship ( $n = 24$ ), and withdrawal ( $n = 24$ ). Detailed explanation of financial COI was found in only 17 retraction notices (9%); 35 notices (18%) included a detailed description of undisclosed or misdislosed employment relationships. Of 48 expressions of concern, errata, corrigendum notices, or undisclosed or concerning COI declarations were the only reason for 16 (33%). The most common additional reasons for a notice were lack of information provided on methods, study data, and process of institutional ethical approval for the published article ( $n = 14$ ), as well as concerns over the completeness of acknowledged contributions ( $n = 5$ ). The median (IQR) time from article publication to retraction was 414 (224-1170) days. Russia and China (38 articles each) were the most common origins of the retracted articles. In logistic regression, the predictive linguistic attributes for retraction notices were higher analytic score (odds ratio [OR], 1.18; 95% CI, 1.07-1.30) and lower clout score (OR, 0.93; 95% CI, 0.90-0.95). The t-SNE visualizing technique found no linear separation between retraction and correction or expression of concern points, indicating that similar language is used across different publication notices.

**Conclusions** Nondisclosed or incorrect disclosure of COI may be a sole reason for article correction, and evident COI may be a sole reason for article retraction. Retraction notices often do not describe in detail the conflicts involved. Higher analytic tone of the retraction notices suggested that editors used more logical reasoning when elaborating retractions. Higher scores on clout for nonretraction notices suggest that it may have been easier for editors to express more confidence in writing these corrections.

## References

1. Oransky I, Marcus A. Why do medical journals keep taking authors at their word? Accessed January 29, 2022. <https://www.statnews.com/2018/09/14/jose-baselga-doctors-industry-conflict/>
2. van der Maaten LJP, Hinton GE. Visualizing data using t-SNE. *J Machine Learning Res.* 2008;9:2579-2605.

<sup>1</sup>Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia, bocinaruzica@gmail.com;

<sup>2</sup>Center for Evidence-Based Medicine, University of Split School of Medicine, Split, Croatia

**Conflict of Interest Disclosures** Ana Marušić is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. No other disclosures were reported.

**Funding/Support** This study was funded by the Croatian Science Foundation “Professionalism in Health–Decision Making in Practice and Research” (ProDem) under grant agreement IP-2019-04-4882.

**Role of the Funder/Sponsor** The funder had no role in the study design, data collection, and analysis; decision to publish; or preparation of the manuscript.

**Acknowledgments** We thank Retraction Watch for providing access to their retraction database.

---

## Characteristics of Articles in Clinical and Translational Sciences Retracted for Reasons Related to the Capture, Management, or Analysis of Data: A Scoping Review

Grace C. Bellinger,<sup>1</sup> Abigail S. Baldridge,<sup>1</sup> Luke V. Rasmussen,<sup>1</sup> Oriana M. Fleming,<sup>1</sup> Eric W. Whitley,<sup>1</sup> Leah J. Welty<sup>1</sup>

**Objective** To characterize clinical and translational science publications retracted for reasons related to the capture, management, or analysis of data to better understand errors that may occur in the research pipeline.

**Design** This scoping review complied with Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines and followed a preregistered protocol. The Retraction Watch database was queried through March 12, 2020.<sup>1,2</sup> Records were eligible for abstract review if they were published between 2010 and 2020, had a subject list containing terms related to clinical and translational science, and were retracted because of concerns with the capture, management, or analysis of data. Abstracts were reviewed in duplicate. A retracted article was eligible for full-text review if abstract review determined it was published in English and

related to clinical and translational sciences. During full-text review, the study team extracted information on number of authors, author attributions, data types and sources, study design, statistical analysis plan, software, and data availability. Research electronic data capture (REDCap) was used for import of publication information from Retraction Watch and data entry throughout the abstract and full-text review processes.<sup>3</sup> A random sample of 5% of articles were reviewed in duplicate. Descriptive analyses were performed using R, version 4.0.1 (<https://www.R-project.org/>).

**Results** Of 21,252 records retrieved from Retraction Watch, 1266 (6%) were eligible for abstract review and 884 (4%) were eligible for full-text review. Of the 884 publications eligible for full-text review, 786 (89%) were available online through Northwestern University’s library system and included in the final analyses. The analytic set included 571 articles (73%) involving human research, 213 reports (27%) of animal research, and 47 systematic reviews or meta-analyses (6%). Few retracted articles described data that were publicly available (67 [9%]) or stated that data were available on request (21 [3%]); most articles contained no statements related to data availability or methods for data capture and management. More than one-third of the retracted articles (300 [38%]) did not specify the statistical analysis software used. The most-used programs were SPSS/PASW (229 [29%]) and GraphPad PRISM (86 [11%]). Statistical software such as Stata (43 [5%]), SAS (38 [5%]), or R/R Studio (29 [4%]) were infrequently reported.

**Conclusions** This scoping review identified more than 800 articles in clinical and translational sciences retracted over 10 years for concerns related to data capture, management, or analysis. The results describe this cohort of retracted articles. Future work will include comparisons with a set of randomly selected publications that have not been retracted but are matched on journal and time frame. Authors can improve the rigor of scientific research by reporting software used and data availability. Publishers, editors, and peer reviewers can contribute to these improvements by advocating for widespread adoption of transparent documentation.

## References

1. Baldridge AS, Welty LJ, Rasmussen L, Whitley E. *Analysis of Retraction Watch Database for Retractions Due to Errors in the Clinical Science Reproducible Research Pipeline: A Scoping Review Protocol*. DigitalHub, Galter Health Sciences Library & Learning Center; 2020. doi:10.18131/g3-pep3-ky27
2. Retraction Watch database. Center for Scientific Integrity. 2018. Accessed March 12, 2020. <http://retractiondatabase.org/>
3. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010

<sup>1</sup>Feinberg School of Medicine, Northwestern University, Chicago, IL, USA, lwealty@northwestern.edu

**Funding/Support** Research reported in this abstract was supported, in part, by the National Institutes of Health's National Center for Advancing Translational Sciences (UL1TR001422).

**Role of the Funder/Sponsor** The National Institutes of Health had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Virtual Posters

### Artificial Intelligence

#### Quality of Reporting of Randomized Clinical Trials in Artificial Intelligence: A Systematic Review

Rehman Siddiqui,<sup>1,2,3</sup> Rida Shahzad,<sup>1</sup> Bushra Ayub<sup>4</sup>

**Objective** The aim of this study was to evaluate the reporting quality of randomized clinical trials (RCTs) of artificial intelligence (AI) in health care from 2015 to 2020 against the Consolidated Standards of Reporting Trials—Artificial Intelligence (CONSORT-AI)<sup>1</sup> guideline.

**Design** In this systematic review, PubMed and Embase databases were searched to identify eligible studies published from 2015 to 2020. Articles were included if AI (defined as AI, machine learning, or deep learning studies) was used as an intervention for a medical condition, if there was evidence of randomization, and if there was a control group in the study. Exclusion criteria were nonrandomized studies, secondary studies, post hoc analyses, if the intervention was not AI, if the target condition was not a medical disease, or if the study pertained to medical education. The included studies were graded by 2 independent reviewers using the CONSORT-AI checklist, which included 43 items. Any disagreements were resolved by consensus following discussion with a senior reviewer. Each item was scored as fully reported, partially reported, or not reported. Irrelevant items were labelled as not applicable. The results were tabulated, and descriptive statistics were reported.

**Results** A total of 939 potential abstracts were screened, from which 73 full-text articles were reviewed for eligibility. Fifteen studies were included in the review. The number of participants ranged from 28 to 1058. Studies pertained to medical fields, including medicine ( $n = 2$ ), psychiatry ( $n = 3$ ), gastroenterology ( $n = 5$ ), cardiology ( $n = 2$ ), ophthalmology ( $n = 1$ ), endocrinology ( $n = 1$ ), and neurology ( $n = 1$ ). Studies were from China ( $n = 6$ ), the United States ( $n = 6$ ), the United Kingdom ( $n = 1$ ), the Netherlands ( $n = 1$ ), and Israel ( $n = 1$ ). Only 3 items of the CONSORT-AI checklist were fully reported in all studies. Five items were not applicable in more

**Disclaimer** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Acknowledgments** We thank the Retraction Watch team as well as Q. Eileen Wafford (Galter Health Sciences Library, Northwestern University).

than 85% of the studies (13 of 15). Twenty percent of the studies (3 of 15) did not report more than 50% of the CONSORT-AI checklist items.

**Conclusions** Reporting quality of RCTs on AI was suboptimal. Because reporting varied in the analyzed RCTs, caution must be exercised when interpreting their outcomes.

#### Reference

1. Liu X, Faes L, Calvert MJ, Denniston AK. Extension of the CONSORT and SPIRIT statements. *Lancet*. 2019;394(10205):1225.

<sup>1</sup>Shahzad Eye Hospital, Karachi, Pakistan, rehman.siddiqui@gmail.com; <sup>2</sup>Department of Ophthalmology and Visual Sciences, Aga Khan University Hospital, Karachi, Pakistan; <sup>3</sup>The Eye Centre, South City Hospital, Karachi, Pakistan; <sup>4</sup>Aga Khan University Hospital, Karachi, Pakistan

**Conflict of Interest Disclosures** None reported.

#### A Machine Learning–Powered Literature Surveillance Approach to Identify High-Quality Studies From PubMed in Disease Areas With Low Volume of Evidence

Patricia L. Kavanagh,<sup>1</sup> Tamara Navarro-Ruan,<sup>2</sup> Peter LaVita,<sup>1</sup> Parrish Rick,<sup>1</sup> Alfonso Iorio<sup>2,3</sup>

**Objective** The DynaMed Systematic Literature Surveillance process surveys a large set of clinical journals most likely to contain high-quality, high-relevance content on treatment, diagnosis, and prognosis across all medical conditions. For many conditions, limited content is retrieved from those journals. Therefore, a machine learning–powered process was designed, implemented, and tested to efficiently and accurately identify relevant articles published across all journals indexed in PubMed.<sup>1–3</sup> This study reports the overall performance of this machine learning–augmented surveillance system.

**Design** Content-based search strategies were developed by a medical librarian. PubMed-retrieved references were probability ranked by a LightGBM machine learning algorithm for likelihood of reporting high-quality, clinically relevant evidence.<sup>1</sup> Top-ranked references were included for screening, stratified by publication date (<18 months or ≥18 months). Clinical experts trained in critical appraisal of the literature manually screened the references and identified those to be used for updating the topic. The following metrics

were used to evaluate the machine learning system: median probability ranking by machine learning of the 15 highest-ranked references, overall and by topic; total and median number of references retrieved by topic; and median position of the first selected reference in the probability-ranked list compared with PubMed reference lists ranked as most recent and best match.

**Results** As of May 2022, results were reviewed for 332 topics. Of 91,009 articles identified, the 8406 (9.2%) with the highest probability ranking were manually screened, and 576 references (6.9%) selected to update 241 topics. The median number of references retrieved by topic was 184 (range, 7-3638). The median probability assigned to the 576 references was 0.047 (range, 0.002-0.996), and the median probability by topic was 0.079 (range, 0.047-0.803). The median position of first selected reference for machine learning was 2 vs 9 for the PubMed most recent strategy and 20 for the PubMed best match strategy. Overall, the median difference in position was 22 for machine learning vs the PubMed most recent strategy and 54.5 for machine learning vs the PubMed best match strategy. The 241 topics were distributed among 29 specialties, with pediatrics and infectious diseases accounting for 27%. The most common article type selected was cohort study (29%).

**Conclusions** This study provides precise estimates of the performance of a regression-based machine learning algorithm in assisting literature surveillance for topics with a low volume of evidence.

## References

1. Abdelkader W, Navarro T, Parrish R, et al. A deep learning approach to refine the identification of high-quality clinical research articles from the biomedical literature: protocol for algorithm development and validation. *JMIR Res Protoc.* 2021;10(11):e29398. doi:10.2196/29398
2. Del Fiol G, Michelson M, Iorio A, Cotoi C, Haynes RB. A deep learning method to automatically identify reports of scientifically rigorous clinical research from the biomedical literature: comparative analytic study. *J Med Internet Res.* 2018;20(6):e10281. doi:10.2196/10281
3. Abdelkader W, Navarro T, Parrish R, et al. Machine learning approaches to retrieve high-quality, clinically relevant evidence from the biomedical literature: systematic review. *JMIR Med Inform.* 2021;9(9):e30401. doi:10.2196/30401

<sup>1</sup>DynaMed, EBSCO Health, Ipswich, MA, USA; <sup>2</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada, iorioa@mcmaster.ca; <sup>3</sup>Department of Medicine, McMaster University, Hamilton, ON, Canada

**Conflict of Interest Disclosures** None reported.

## Bias

### Development of a New Risk of Bias Tool for Network Meta-analysis (RoB NMA Tool)

Carole Lunny,<sup>1,2</sup> Areti-Angeliki Veroniki,<sup>1,3</sup> Brian Hutton,<sup>3,4,5</sup> Ian R. White,<sup>6</sup> Julian P. T. Higgins,<sup>7,8,9</sup> James M. Wright,<sup>10</sup> Sofia Dias,<sup>11</sup> Penny Whiting,<sup>12</sup> Andrea C. Tricco<sup>1,13,14</sup>

**Objective** Researcher and stakeholder interaction in the development of new tools to inform evidence-based medicine is a key factor associated with the impact such tools can have. Currently, there no risk of bias (RoB) tools to assess reviews including network meta-analyses (NMAs; **Table 76**). The objectives of this research were to identify items for potential inclusion in the tool through a methodological systematic review, conduct a Delphi survey, and conduct a stakeholder survey.

**Design** An international steering committee developed a protocol following the methods by Whiting et al<sup>1</sup> for tool development and made conceptual decisions about the tool's structure. Tools, articles, and editorial standards presenting items related to bias, reporting, or quality in NMAs were included. General systematic review items were excluded. Experts for the Delphi survey were identified using a purposive sampling. Respondents were asked to rate whether items should be included. All agreed-upon items (defined as 70% agreement) and additional or aggregated items were included in a second round of the survey. The stakeholder survey contained 22 questions and was disseminated anonymously through social media and professional networks.

**Results** The search returned 3599 citations, from which 59 articles were included, yielding 99 items.<sup>2</sup> Of these, 22 items were deemed eligible and were entered into a Delphi survey in which 26 respondents completed round 1 and 22 completed round 2.<sup>3</sup> Seven items did not reach consensus in round 2 of the Delphi survey. After further refinement by the committee, 16 items were worded as signaling questions and categorized into 3 domains in the tool. An elaboration and explanation document was drafted. A total of 298 stakeholders participated in the survey; 75% indicated that their organization produced NMAs, and 78% showed high interest in the tool.<sup>3</sup> Most stakeholders (84%) who responded to the survey reported they would use the tool to assess an NMA if they had received adequate training. Most stakeholders and Delphi panelists preferred a tool to assess both bias in NMA results and authors' conclusions. After examining the results of these studies, the committee recommended that the tool be used with the ROBIS tool for assessing biases in systematic reviews using a domain-based structure and to assess both NMA results and authors' conclusions. Response bias in this sample was a major limitation, as stakeholders and Delphi panelists working in higher-income countries were more represented.

**Table 76. Instruments to Aid in Systematic Review Conduct or to Assess the Reporting or Methodological Quality of a Review<sup>a</sup>**

Tool purpose	Examples of instruments	Description of an example tool	Available tool for reviews with NMA
Guidance for conducting systematic reviews	MECIR	Detailed guidance for the conduct of systematic reviews of interventions, diagnostic test accuracy, individual patient data, public health, and health promotion.	No
Assess the quality of published reviews	AMSTAR-2, OQAQ	AMSTAR-2 is a critical appraisal tool to assess the conduct of intervention reviews, including RCTs.	No
Assess the risk of bias of published reviews	ROBIS	ROBIS is a tool for assessing the risk of bias in reviews. It is aimed at 4 broad categories of reviews, mainly within health care settings: interventions, diagnosis, prognosis, and etiology.	In process (RoB NMA tool)
Assess the certainty in a body of evidence	GRADE	The GRADE working group defined the certainty of a body of evidence as the extent to which one can be confident that a pooled effect estimate is close to the true effect of the intervention. Five domains assessed: risk of bias, inconsistency, indirectness, imprecision, and publication bias.	GRADE-NMA, CINEMA, threshold method
Guidelines for the complete reporting of published reviews	PRISMA update	PRISMA focuses on the reporting of already published reviews evaluating RCTs of interventions. PRISMA can determine whether a review is well described and transparently reported.	PRISMA-NMA

Abbreviations: AMSTAR-2, A Measurement Tool to Assess Systematic Reviews 2; CINEMA, Confidence in Network Meta-analysis; GRADE, Grading of Recommendations Assessment, Development and Evaluation; MECIR, Methodological Expectations of Cochrane Intervention Reviews; NMA, network meta-analysis; OQAQ, overview quality assessment questionnaire; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT, randomized clinical trial; RoB, risk of bias; ROBIS, Risk of Bias in Systematic Reviews.

<sup>a</sup>Currently, several checklists exist for critically appraising reviews with NMA (eg, the PRISMA statement extension for reviews incorporating NMA, the National Institute for Health and Care Excellence Decision Support Unit checklist for assessing reporting quality, the International Society for Pharmacoeconomics and Outcomes Research checklist for assessing credibility and applicability; and Dias 2018 for assessing validity). These tools were designed with different purposes: some for assessing reporting quality in reviews with NMA, some for assessing conduct, applicability, or validity. These review-level tools are not to be confused with tools to assess the individual primary studies included in systematic reviews (eg, Cochrane RoB tool for randomized trials). This table compares the different tools, indicating whether an equivalent tool for reviews including NMAs exists.

**Conclusions** These studies inform the development of the first tool to assess RoB in NMAs. In the future, the tool will be pilot tested in different user groups.

## References

1. Whiting P, Wolff R, Mallett S, Simera I, Savović J. A proposed framework for developing quality assessment tools. *Syst Rev*. 2017;6(1):1-9. doi:10.1186/s13643-017-0604-6
2. Lunny C, Veroniki AA, Hutton B, et al. Methodological review to develop a list of bias items used to assess reviews incorporating network meta-analysis. Submitted to *Research Synthesis Methods*, 2022.
3. Lunny C, Tricco AC, Veroniki AA, et al. Stakeholder opinions on the structure and development of a new risk of bias tool to assess systematic reviews with network meta-analysis (RoB NMA tool): a cross-sectional survey. *Research Square*. Preprint posted online February 22, 2022. doi:10.21203/rs.3.rs-1324758/v1

<sup>1</sup>Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada, carole.lunny@ubc.ca; <sup>2</sup>Cochrane Hypertension Review Group, Therapeutics Initiative, The University of British Columbia, Vancouver, BC, Canada; <sup>3</sup>Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada; <sup>4</sup>Ottawa Hospital Research Institute, Ottawa, ON, Canada; <sup>5</sup>School of Epidemiology and Public Health, Ottawa University, Ottawa, ON, Canada; <sup>6</sup>Medical Research Council Clinical Trials Unit at University College London, London, UK; <sup>7</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; <sup>8</sup>National Institute for Health and Care Research Bristol Biomedical Research Centre at University Hospitals Bristol and Weston National Health Service Foundation Trust and the University of Bristol, Bristol, UK; <sup>9</sup>National Institute for Health and Care Research Applied Research Collaboration West (ARC West) at University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK; <sup>10</sup>Cochrane Hypertension Review Group and the Therapeutics Initiative, University of British Columbia, Vancouver, BC, Canada; <sup>11</sup>Centre for Reviews and Dissemination, University of York, York, UK; <sup>12</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; <sup>13</sup>Dalla Lana School of Public Health & Institute of Health Policy, Management, and

Evaluation, University of Toronto, Toronto, ON, Canada; <sup>14</sup>Queen's Collaboration for Health Care Quality Joanna Briggs Institute Centre of Excellence, Queen's University, Kingston, ON, Canada

**Conflict of Interest Disclosures** Brian Hutton has previously received honoraria from Eversana Incorporated for the provision of methodologic advice related to the conduct of systematic reviews and meta-analyses. Ian R. White was supported by the Medical Research Council (programme MC\_UU\_00004/06). Julian P. T. Higgins is a National Institute for Health Research (NIHR) Senior Investigator (NF-SI-0617-10145), is supported by NIHR Bristol Biomedical Research Centre at University Hospitals Bristol and Weston National Health Service (NHS) Foundation Trust and the University of Bristol, is supported by the NIHR Applied Research Collaboration West at University Hospitals Bristol and Weston NHS Foundation Trust, and is a member of the Medical Research Council (MRC) Integrative Epidemiology Unit at the University of Bristol. Andrea C. Tricco holds a Tier 2 Canada Research Chair in Knowledge Synthesis. No other disclosures were reported.

**Funding/Support** This study was supported through a 2020 Canadian Institutes of Health Research Project Grant (2021-2024; ID 174998).

**Role of the Funder/Sponsor** The funder had no role in the submitted work.

**Acknowledgments** We acknowledge the panelists in the Delphi study: Simon Turner, Dimitris Mavridis, Virginia Chiocchia, Ian Shrier, Adriani Nikolakopoulou, Dan Jackson, Richard Riley, Becky Turner, Bruno Roza da Costa, Petros Pechlivanoglou, Steve Kanters, Ferran Catala-Lopez, Tianjing Li, Matt Page, Chris Cameron, Anna Chaimani, Kerry Dwan, Audrey Beliveau, Kristian Thorlund, Jonathan Sterne, Cinzia del Giovane, Guido Schwarzer, Jo McKenzie, Lehana Thabane, Theodoros Papakonstantinou, Isabelle Boutron, Sharon Straus, and Jenn Watt. We also thank the participants of the stakeholder survey, who were anonymous.

**Additional Information** The views expressed in this article are those of the authors and do not necessarily represent those of the NHS, the NIHR, the MRC, or the Department of Health and Social Care.

## Bias in Meta-analysis Estimates Associated With Varying Quality of Patient-Reported Outcome Measures in Orthopedics

Joel J. Gagnier,<sup>1,2</sup> Jianyu Lai<sup>2</sup>

**Objective** A previous study revealed that patient-reported outcome measures (PROMs) with poor or unknown psychometric properties were associated with higher estimates of treatment effect in clinical trials of rotator cuff diseases.<sup>1</sup> This study assessed the variations in meta-analysis estimates in orthopedics associated with varying quality of PROMs and hypothesized an average higher estimate in PROMs with unknown and poor psychometric properties.

**Design** Meta-analyses were identified from 5 databases from inception through October 16, 2017. PROM scores were derived from a prior publication that comprehensively assessed the quality of these instruments (higher scores were better quality).<sup>2</sup> Standardized mean difference (effect size) or mean difference of change in PROM scores (from before treatment to after treatment) between different treatment types were extracted or calculated for each study. For those studies that did not report standardized results, change scores were divided by the SD for standardization. The SD was imputed in some cases from SEs and CIs. A mixed-effects regression analysis was done, with all standardized change scores as dependent variables and other data as independent variables (PROM overall quality score, number of studies, total sample size across included studies, and average follow-up), controlling for the grouping variable meta-analysis (in which multiple estimates were calculated for several PROMs from within the same meta-analysis). A sensitivity analysis was done excluding meta-analytic estimates for mixed interventions. Increases in  $\beta$  coefficients indicate effect size change for each unit increase in PROM quality.

**Results** A total of 249 unduplicated meta-analyses on rotator cuff disease were reviewed with 47 being included, with 6 different PROMs included, and several meta-analyses included mixed outcomes with several PROMs being combined. Reviews were excluded (202) primarily because one of the PROMs of interest was not used. The  $\beta$  coefficient for PROM quality and the pooled effect size estimates was  $-0.012$  (95% CI,  $-0.049$  to  $0.025$ ;  $P = .53$ ) before and after controlling for several covariates (**Table 77**). In the sensitivity analysis, after removing meta-analyses with mixed PROMs in the pooled effect size estimates, the  $\beta$  coefficient for PROM quality and pooled effect estimates was  $-0.013$  (95% CI,  $-0.034$  to  $0.007$ ;  $P = .19$ ); this finding was not statistically significant.

**Conclusions** In estimating the percentage of bias, this study found that pooled effect size estimates across PROMs of poor quality inflate effect estimates by approximately 10% (the ratio of the  $\beta$  estimate and meta-analytic estimate). This magnitude of effect size is not statistically significant, but

**Table 77. Mixed-Effects Meta-regression Findings for Changes From Baseline**

Variable	Standardized mean difference, $\beta$ (95% CI)	P value
<b>Initial analysis<sup>a</sup></b>		
PROM quality	$-0.012$ ( $-0.049$ to $0.025$ )	.53
Sample size (No. of patients)	$-0.0001$ ( $-0.001$ to $0.001$ )	.79
No. of included studies	$0.001$ ( $-0.013$ to $0.017$ )	.86
<b>Sensitivity analysis<sup>b</sup></b>		
PROM quality	$-0.013$ ( $-0.034$ to $0.007$ )	.19
Sample size (No. of patients)	$0.0001$ ( $-0.001$ to $0.001$ )	.85
No. of included studies	$0.002$ ( $-0.005$ to $0.010$ )	.50

Abbreviation: PROM, patient-reported outcome measure.

<sup>a</sup>This analysis included 44 outcomes across 18 meta-analyses, 6 of which had pooled estimates across different PROMs.

<sup>b</sup>This analysis included 28 outcomes across 12 meta-analyses, excluding those meta-analyses that had pooled data across differing PROMs.

larger methodologic studies may be warranted to confirm clinical significance.

## References

1. Gagnier JJ, Johnston BC. Poor quality patient reported outcome measures bias effect estimates in orthopaedic randomized studies. *J Clin Epidemiol*. 2019;116:36-38.
2. Huang S, Grant J, Miller B, Mirza FM, Gagnier JJ. A systematic review of psychometric properties of patient reported outcome instruments for use in patients with rotator cuff disease. *Am J Sports Med*. 2015;43(10):2572-2582.

<sup>1</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA, jgagnier@umich.edu; <sup>2</sup>Department of Orthopaedic Surgery, University of Michigan Health System, Ann Arbor, MI, USA

**Conflict of Interest Disclosures** None reported.

**Additional Information** Jianyu Lai is a co-corresponding author.

## Development and Pilot Test of Risk of Bias Assessment Tool for Use in Peer Review

Brian S. Alper,<sup>1</sup> Joanne Dehnboestel,<sup>1</sup> Khalid Shahin,<sup>1</sup> Amy Price,<sup>2</sup> for the COVID-19 Knowledge Accelerator (COKA) Initiative

**Objective** Peer review of clinical research should include risk of bias assessment, but it is often limited and not systematic. The Risk of Bias Assessment Tool (RoBAT) is an open web-based tool in which the user can select types of bias from a hierarchical list and record their assessment to help document a comprehensive risk of bias assessment for any scientific study. The RoBAT Usability Research Pilot Study was done to inform development of the tool and measures of its initial effectiveness, efficiency, and satisfaction.

**Design** Participants were recruited through email distribution lists of COVID-19 Knowledge Accelerator, Guidelines International Network, Health Level Seven International, Healthcare Information For All, and

International Society for Evidence-based Health Care (January 28-31, 2022). An online platform enabled participants to attest to meeting eligibility criteria (experience or education regarding risk of bias assessment and willingness to complete the study online), consent to participate, and complete usability evaluation reports for each assessment they attempted. Participants were encouraged to complete assessments at least 3 times (including 1 without RoBAT, 1 with RoBAT, and 1 with RoBAT after study-generated improvements). Surveys for each assessment included whether the task was completed (participant defined), time-on-task (participant reported as number of minutes), perceived ease of task (5-point scale), perceived ease of use of the tool (5-point scale), and open-ended questions for likes, dislikes, and suggested improvements. End-of-study surveys included the System Usability Scale<sup>1</sup> and perceived usefulness for peer review support. Study enrollment closed after more than 5 participants used RoBAT, the number needed to detect 80% to 85% of issues for the initial discovery of usability problems.<sup>2,3</sup>

**Results** A total of 18 participants were enrolled in the study and 10 completed 32 risk of bias assessment attempts. Task completion was achieved for 6 of 7 (86%) attempts without RoBAT, 13 of 17 (77%) attempts with the initial version of RoBAT, and 5 of 8 (63%) attempts with the revised version of RoBAT. Median (range) time on task was 34.5 (5-60) minutes without RoBAT, 30 (5-90) minutes with the initial version of RoBAT, and 20 (14-120) minutes with the revised version of RoBAT. The most common suggestions for improvements were to add instructions and to facilitate rapid selection of recognized terms. Data were too limited to establish a pattern for ease-of-use ratings from no to initial to revised RoBAT use. Five of 9 participants reported they would likely use RoBAT for peer review support.

**Conclusions** The pilot study provided preliminary evidence of efficiency and satisfaction with RoBAT and demonstrated the feasibility of rapid online research development and implementation. Subsequent developments could test the tool's usefulness for systematic reviewers and journal editors and integration with editorial systems.

## References

1. Lewis JR. The System Usability Scale: past, present, and future. *Int J Hum Comput Interact*. 2018;34(7):577-590. doi: 10.1080/10447318.2018.1455307
  2. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. In: *CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. Association for Computing Machinery; 1993:206-213. doi:10.1145/169059.169166
  3. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? *Hum Factors*. 1992;34(4):457-468. doi:10.1177/001872089203400407
- <sup>1</sup>Scientific Knowledge Accelerator Foundation, Computable Publishing LLC, Ipswich, MA, USA, balper@computablepublishing.com; <sup>2</sup>Anesthesia Informatics and Media Lab, Stanford University School of Medicine, Palo Alto, CA, USA
- Conflict of Interest Disclosures** Brian S. Alper owns Computable Publishing LLC. Joanne Dehnbozel and Khalid Shahin are employees of Computable Publishing LLC. No other disclosures were reported.
- ## Bias, Publication
- ### Assessment of Gender Balance in the Editorial Activities of a Researcher-Led Journal
- Tal Seidel Malkinson,<sup>1</sup> Devin B. Terhune,<sup>2</sup> Mathew Kollamkulam,<sup>3</sup> Maria J. Guerreiro,<sup>4</sup> Dani S. Bassett,<sup>5,6,7,8,9,10</sup> Tamar R. Makin<sup>3</sup>
- Objective** Editorial decision-making is a fundamental element of the scientific enterprise, with critical implications for career advancement. Despite repeated calls for making deliberate efforts to incorporate gender diversity into editorial board structures, gender disproportions remain pervasive.<sup>1,2</sup> Gender parity in the contributions to editorial decisions at various stages of the publication process was examined, based on analytics collected by the biomedical researcher-led journal *eLife*.
- Design** Data accumulated by *eLife*'s platform from 2017 to 2019 were organized into 2 data sets. The reviewing editor (RE) data set included anonymous information on the engagement of individual REs ( $n = 1201$ ) in the editorial process, with a binary gender assigned based on the editor's name and gender expression. REs were consulted by senior editors at the initial assessment stage, and an RE was chosen to handle the full review process for the selected manuscripts. The manuscript data set included the outcome of submitted manuscripts ( $n = 24,056$ ) in each submission stage, the assigned gender of the REs suggested by the authors, the assigned gender of the handling RE, and the assigned gender of the appointed senior editor. Owing to nonnormal distributions in the data, 2-tailed nonparametric tests were used, including (1) binomial tests and  $N - 1 \chi^2$  proportion comparison tests, (2) contingency table analysis, (3) a permutation-based Welch independent *t* test, and (4) equivalent bayesian analyses when significance was close to  $P < .05$ .
- Results** Despite efforts to increase women representation, the board of REs was predominantly male (833 [69.4%]). Authors suggested fewer women as REs, even after correcting for men overrepresentation (29.08% women vs 30.6% men;  $\chi^2_1 = 11.65$ ;  $P = .001$ ; Cohen  $h = 0.90$ ). Although women editors were proportionally involved in the initial manuscript assessment (mean [SD] number of assessment requests per month, 2.40 [1.44] women REs vs 2.41 [1.51] men REs;  $t_{809.7} = 0.11$ ;  $P = .92$ ), they were underengaged in the full review process (mean [SD] number of full submissions per month, 0.40 [0.32] women REs vs 0.44 [0.37] men REs;  $t_{869.8} = 2.22$ ;  $P = .03$ ; Hedges  $g = 0.13$ ). Gender homophily in manuscript

assignment was found, such that senior editors overengaged same-gender REs ( $\chi^2_1 = 224.55$ ;  $P < .001$ ; contingency coefficient of 0.186) (Figure 28). This tendency was stronger in more gender-balanced scientific disciplines (eg, in developmental biology, with 56.9% of manuscripts handled by men REs;  $r = -0.47$ ;  $P = .05$ ; Bayes factor<sub>10</sub> = 1.77).

**Conclusions** Together, the findings confirm that gender disparities exist along the editorial process and suggest that merely increasing the proportion of women members might not be sufficient to eliminate this bias.

## References

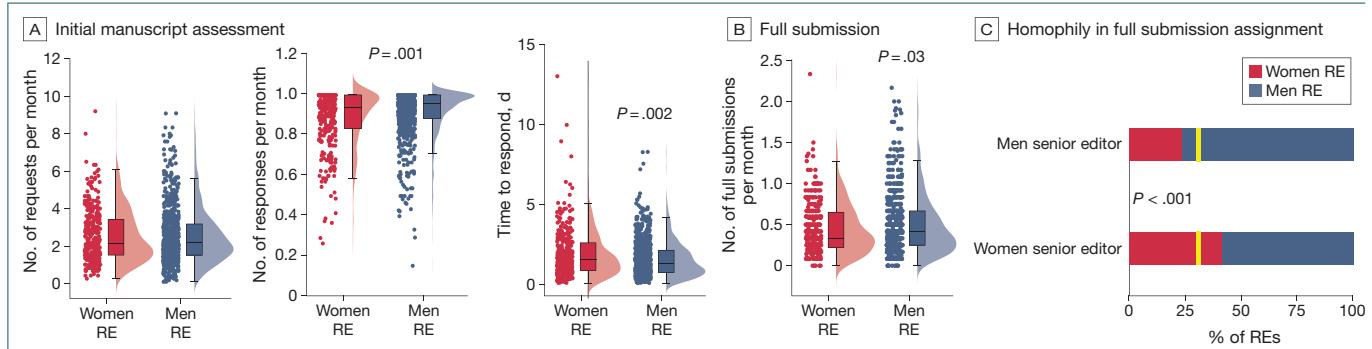
1. Helmer M, Schottendorf M, Neef A, Battaglia D. Gender bias in scholarly peer review. *eLife*. 2017;6:e21718. doi:10.7554/eLife.21718
2. Palser ER, Lazerwitz M, Fotopoulou A. Gender and geographical disparity in editorial boards of journals in psychology and neuroscience. *Nat Neurosci*. 2022;25(3):272-279. Medline:35190729 doi:10.1038/s41593-022-01012-w

<sup>1</sup>Sorbonne Université, Institut du Cerveau–Paris Brain Institute–ICM, INSERM, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France, tal.seidel@mail.huji.ac.il; <sup>2</sup>Department of Psychology, Goldsmiths, University of London, London, UK; <sup>3</sup>Institute of Cognitive Neuroscience, University College London, London, UK; <sup>4</sup>eLife Sciences Publications Ltd, Cambridge, UK; <sup>5</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA; <sup>6</sup>Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA; <sup>7</sup>Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA, USA; <sup>8</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA; <sup>9</sup>Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA; <sup>10</sup>Santa Fe Institute, Santa Fe, NM, USA

**Conflict of Interest Disclosures** Tamar R. Makin is a senior editor for *eLife*. Maria J. Guerreiro is part of the executive staff team of *eLife*.

**Funding/Support** Tal Seidel Malkinson was funded by Agence Nationale de la Recherche grant ANR-16-CE37-0005; Tamar R. Makin was funded by Wellcome Trust Senior Research Fellowship grant 215575/Z/19/Z and ERC Starting Grant 715022 EmbodiedTech.

**Figure 28. Gender Imbalance in Editorial Activities**



Gender differences in initial assessment: Senior editors equally engaged women and men reviewing editors (REs) in the initial consultation. Women REs responded slightly less to senior editor's initial consultation requests, and they took longer to respond than did men REs. B, Men REs handled slightly more full submissions per month than did women REs. C, Compared with gender base rates of manuscript assignment (yellow lines; base rate of RE manuscript assignment: 30.0% women REs; 70.0% men REs), men senior editor assigned significantly more manuscripts to men REs (4815 of 6289 [76.6%]), and women senior editors assigned significantly more manuscripts to women REs (2604 of 6289 [41.4%]).

**Role of Funder/Sponsor** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; or decision to submit the abstract for presentation.

**Additional Contributions** We thank the *eLife* editorial team for facilitating this study at all stages and, in particular, Andy Collings and Daniel Ecer for help with preparing the data set; Jennifer Raymond, Chris Baker, and Jon Roiser for discussions about the data set; and Michael B. Eisen, Tim Behrens, and *eLife*'s senior leadership for feedback and ongoing support of this study.

**Additional Information** Maria J. Guerreiro is a co-corresponding author.

## Spin in Randomized Clinical Trials of Top Medical Journals

Karina Raygoza-Cortez,<sup>1</sup> Francisco Barrera,<sup>1</sup> Mariano García-Campa,<sup>1</sup> Sofía Mariño-Velasco,<sup>1</sup> Melissa Sáenz-Flores,<sup>1</sup> Patricia Castillo-Morales,<sup>1</sup> Miguel Zambrano-Lucio,<sup>1</sup> Augusto Gamboa-Alonso,<sup>1</sup> Amanda Rojo-Garza,<sup>1</sup> José Gerardo González-González,<sup>1</sup> Rene Rodríguez-Gutiérrez<sup>1,2</sup>

**Objective** To analyze the proportion of spin strategies present in randomized clinical trials (RCTs) published in high-impact journals appraised at low risk of bias.

**Design** For this cross-sectional study, a comprehensive search was made in Ovid MEDLINE, retrieving all RCTs published in the 10 highest-impact medical journals in medicine and surgery (by 2018 Journal Citation Reports impact factor) from January 1, 2018, to February 28, 2020. Stratified sampling was then performed of 150 articles, adjusted by journal. Then, the risk of bias of each RCT was assessed independently and in duplicate using the Cochrane risk-of-bias tool. For each included article, an adaptation of the classification scheme created by Boutron et al<sup>1</sup> for RCTs was used to identify spin strategies in every section of the article.

**Results** A total of 46 RCTs at low risk of bias were appraised using the spin classification scheme. From those included in the analysis, 25 studies were published in internal or general medicine journals and 21 in surgery journals. The number of patients included in the RCTs ranged from 20 to 12,000. The

most spin was identified in the discussion and conclusion sections, with the most common strategies being focusing on statistically significant secondary outcomes (4 [8.7%]), ruling out adverse events (4 [8.7%]), acknowledging statistically nonsignificant results for the primary outcome but emphasizing the beneficial effect of treatment (4 [8.7%]), and focusing only on statistically significant results (5 [10.9%]). In subgroup analyses, a total of 38 spin strategies were identified in studies with nonsignificant primary outcomes vs 11 in studies with significant results. Assessment according to specialty identified 35 spin strategies in surgery journals and 14 in internal medicine journals.

**Conclusions** Overall, the use of spin strategies was found to be more common in studies with a statistically nonsignificant primary outcome vs those with a significant primary outcome. This study highlights the need to raise awareness about the use of spin, even in studies published in high-impact journals and at low risk of bias, as the use of these strategies can affect readers' decision-making.

## Reference

- Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303(20):2058-2064. doi:10.1001/jama.2010.651

<sup>1</sup>Plataforma INVEST UANL-KER Unit Mayo Clinic (Knowledge and Evaluation Research [KER] Unit Mexico), School of Medicine, Universidad Autonoma de Nuevo Leon, Monterrey, Mexico, rodriguezgutierrez.rene@mayo.edu; <sup>2</sup>Knowledge and Evaluation Research (KER) Unit, Mayo Clinic, Rochester, MN, USA

**Conflict of Interest Disclosures** None reported.

## Analysis of Reporting Bias in Published and Unpublished Trials of Extended-Release Alprazolam for Panic Disorder

Rosa Y. Ahn-Horst,<sup>1,2,3</sup> Erick H. Turner<sup>4,5</sup>

**Objective** According to meta-analyses and practice guidelines, benzodiazepines are effective in the treatment of panic disorder.<sup>1,2</sup> However, to date, no meta-analyses have incorporated data from unpublished trials. Among all benzodiazepines, alprazolam is the most widely prescribed and has the highest frequency of nonmedical use, abuse, and related harms in the US.<sup>3</sup> This study examined reporting bias with the extended-release (XR) formulation of alprazolam by comparing its efficacy for panic disorder using trial results from the published literature and the US Food and Drug Administration (FDA).

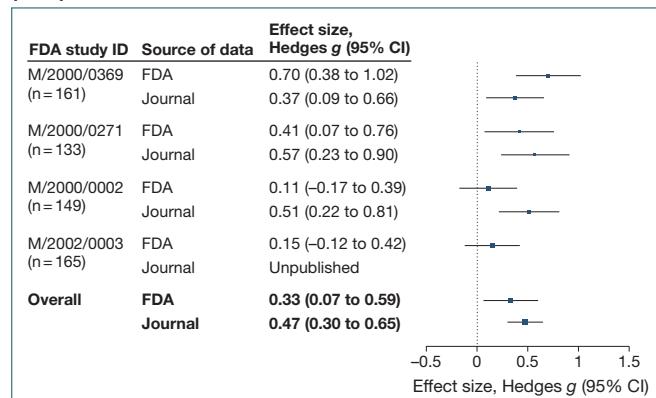
**Design** There was no protocol for this study, and it was not registered. Medical and statistical reviews for alprazolam XR were downloaded from Drugs@FDA (<https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm>); all phase 2 and 3 randomized, double-blind, placebo-controlled efficacy trials were identified; summary statistics on 5 primary outcome measures were extracted; and the FDA's regulatory decision as to whether, for purposes of approval, the trial provided evidence of efficacy (statistical superiority to placebo

on all primary outcomes) was also extracted. For each FDA-registered trial, the published literature was searched for matching publications using PubMed, bibliographies of review articles, and Google Scholar. The best match between FDA-registered trials and publications was based on drug name, comparator, dosage groups, sample size, duration, and investigator name. Summary data on the drug-placebo comparison and whether the publication conveyed that the drug was effective were extracted. Two meta-analyses were conducted—one based on the FDA review and the other based on the published literature—and their effect sizes were compared. Reporting bias was examined by comparing the following: (1) overall trial results (positive or not) according to the FDA vs corresponding publications and (2) effect size (Hedges'  $g$ ) using FDA data vs published data. Risk of bias was not assessed because the objective was not to assess bias in trial methods (internal validity) but rather bias in results reporting.

**Results** The FDA review showed that 5 trials were conducted, only 1 of which (20%) was positive and published (as positive). The remaining 4 studies failed to demonstrate efficacy. Of those, 2 were not published; for the other 2, the articles selectively reported positive, nonprimary, or post hoc outcomes. Thus, according to the published literature, 3 of 3 trials (100%) appeared to show positive results. Alprazolam's overall effect size calculated using FDA data was 0.33 (95% CI, 0.07-0.59), while that based on published trial data was 0.47 (95% CI, 0.30-0.65), an increase of 0.14, or 42% (**Figure 29**).

**Conclusions** According to the results of this analysis, reporting bias has inflated the apparent efficacy of alprazolam XR, as previously found with other drug classes. Because this inflation alters the risk-benefit ratio, clinicians may wish to reconsider their prescribing practices with respect to this benzodiazepine. This study highlights the value of regulatory data to public health.

**Figure 29. Efficacy of Extended-Release Alprazolam for Panic Disorder Based on Data From US Food and Drug Administration (FDA) vs Published Literature**



The forest plot compares effect sizes (Hedges'  $g$  with 95% CI) for trial data as reported in FDA reviews vs corresponding journal publications. One nonsignificant trial (M/2002/0032) was excluded because of a lack of summary statistics reported in the FDA review. The overall effect size based on journal publications exceeded the effect size based on FDA data by 0.14 or 42%.

## References

1. Stein MB, Goin MK, Pollack MH, et al. *Practice Guideline for the Treatment of Patients With Panic Disorder*. American Psychiatric Association; 2010.
2. Wilkinson G, Balestrieri M, Ruggeri M, Bellantuono C. Meta-analysis of double-blind placebo-controlled trials of antidepressants and benzodiazepines for patients with panic disorders. *Psychol Med*. 1991;21(4):991-998. doi:10.1017/S0033291700029986
3. Ait-Daoud N, Hamby AS, Sharma S, Blevins D. A review of alprazolam use, misuse, and withdrawal. *J Addict Med*. 2018;12(1):4-10. doi:10.1097/adm.0000000000000350

<sup>1</sup>Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA, rahn-horst@partners.org; rosa.ahn.horst@gmail.com; <sup>2</sup>Department of Psychiatry, McLean Hospital, Belmont, MA, USA; <sup>3</sup>Department of Psychiatry, Harvard Medical School, Boston, MA, USA; <sup>4</sup>Behavioral Health and Neurosciences Division, Veterans Affairs Portland Health Care System, Portland, OR, USA; <sup>5</sup>Department of Psychiatry, Oregon Health & Science University, Portland, OR, USA

**Conflict of Interest Disclosures** Rosa Y. Ahn-Horst reported no conflicts of interest. Erick H. Turner previously worked as a medical officer for the US Food and Drug Administration and reviewed applications submitted by pharmaceutical companies to determine whether they should be approved for US marketing. He has no financial interest in any pharmaceutical products, approved or otherwise.

## Bibliometrics, Informatics, and Scientometrics

### Comparison of Bibliometrics of Leading Open Access Chinese Journals With Leading Non-Chinese Journals in Science, Technology, and Medicine

Fang Lei,<sup>1</sup> Min Dong,<sup>1</sup> Xuemei Liu<sup>1</sup>

**Objective** This research aimed to perform a comparison of basic bibliometrics of leading Chinese open access (OA) journals from the Excellence Action Plan for Chinese Science, Technology, and Medicine Journals. All journals used English as the publication language.

**Design** This was a cross-sectional investigation reported following the STROBE checklist for conference abstracts. Journal data were extracted from the Web of Science (WoS) and Directory of Open Access Journals databases. All searches were performed in November 2021, and the Journal Citation Reports 2020 impact factor release was considered the latest available data set that allowed for retrieval of relative data. The top 2 international OA journals in the first quartile of the journal impact factor rank based on WoS categories for each leading Chinese OA journal were chosen as control journals. Bibliometric indicators, such as the volume of published documents, citations, and Eigenfactor score (based on items published between 2018 and 2020),

were compared between leading Chinese OA journals and control journals with nonparametric tests.

**Results** A total of 14 of 22 (63.64%) leading Chinese journals were OA journals. For 2 OA journals, there was no corresponding international OA journal; for another 2 OA journals, there was only 1 corresponding OA journal. Therefore, there were 22 control journals. The median (IQR) number of published documents (articles and reviews) per leading Chinese OA journal was less than that of the control journals (302 [243-352] vs 715 [390-2010];  $P = .004$ ); the median (IQR) citations per document (13 [9-26] vs 14 [9-25];  $P = .99$ ) and percentage of cited documents (94.81% [93.63%-98.24%] vs 96.96% [92.64%-98.14%];  $P = .76$ ) were not statistically different. The median (IQR) normalized citation impact of the 2 groups was not significantly different (2.008 [1.221-2.933] vs 2.003 [1.353-3.093];  $P = .97$ ), but the median (IQR) Eigenfactor score, which took into account both the number of citations and the academic influence of citing journals, of leading Chinese OA journals was lower than that of control journals (0.005 [0.003-0.007] vs 0.013 [0.005-0.052];  $P = .02$ ). There were fewer top 1% documents in leading Chinese OA journals (median [IQR], 10 [5-19] vs 34 [11-65];  $P = .02$ ); however, the percentage of top 1% documents was not statistically different (median [IQR], 4.54% [1.18%-5.82%] vs 3.84% [1.86%-6.86%];  $P = .94$ ). The average citations per top 1% document of leading Chinese OA journals were more than that of control journals, but the difference was not statistically significant (median [IQR], 103 [52-134] vs 72 [58-105];  $P = .26$ ).

**Conclusions** Compared with top international OA journals, Chinese leading OA journals published fewer articles with similar average citations, but they were less cited by highly cited journals.

<sup>1</sup>West China Periodicals Press of West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China, liuxuemei@wchscu.cn

**Conflict of Interest Disclosures** None reported.

### A Systematic Review of Medical and Clinical Research Landscapes in Primary Medical Care in Malaysia

Boon-How Chew,<sup>1</sup> Shaun Wen Huey Lee,<sup>2</sup> Lim Poh Ying,<sup>3</sup> Soo Huat Teoh,<sup>4</sup> Aneesa Abdul Rashid,<sup>1</sup> Navin Kumar Devaraj,<sup>1</sup> Adibah Hanim Ismail Daud,<sup>1</sup> Abdul Hadi Abdul Manap,<sup>1</sup> Fadzilah Mohamad,<sup>1</sup> Aaron Fernandez,<sup>5</sup> Hanifatiyah Aliy,<sup>1</sup> Puteri Shanaz Jahn Kassim,<sup>1</sup> Nurainul Hana Shamsuddin,<sup>1</sup> Noraina Muhamad Zakuhan,<sup>6</sup> Akiza Roswati Abdullah,<sup>7</sup> Indah S. Widyahening<sup>8</sup>

**Objective** This systematic review aimed to describe the characteristics of clinical and biomedical research in Malaysia.

**Design** A search was conducted on PubMed, Embase, CINAHL, PsycINFO, and MyMedR (<http://mymedr.afpm.org.my/>) for published clinical and biomedical research in

primary care settings from 1962 to 2017 by Malaysian authors in a Malaysian institution. Studies found were independently screened by a team of reviewers and information was extracted.<sup>1,2</sup> In phase 1, the characteristics of the research and profiles of the researchers and journals in which they were published were reported descriptively. In phase 2, the quality of studies included in phase 1 will be assessed using a newly developed tool to ascertain risk of bias. Longitudinal trends of the research characteristics, health conditions studied (International Classification of Primary Care), and settings, among other characteristics, were explored. No synthesis of results was conducted as no effect estimates were available to be pooled.

**Results** Of 4513 articles, 1078 were included in this qualitative synthesis and 790 with complete data were analyzed. Clinical studies (81.9%), primary research (81.1%), and quantitative studies (74.2%), consisting mostly of prevalence studies (67.7%) by cross-sectional sampling (70.4%), were predominant. The number of studies increased (**Figure 30**) and the number of characteristics also increased after year 2000. Researchers from family medicine (39.3%) and public health (15.2%) specialties were the main contributors to the articles (**Figure 30**, A). Most of the corresponding authors had a master of medicine degree (46.5%) compared with a doctor of philosophy (PhD) (24.4%) or doctor of medicine (MD) (5.2%) degree. Researchers with PhD and MD degrees were more likely to conduct interventional studies compared with those with master's degrees (8.0% vs 6.4% of studies;  $\chi^2 = 54.26$ ;  $P = .03$ ). Publications were mainly original research (82.8%) in international (48.6%) or local (35.0%) journals and were evenly distributed between multidisciplinary (51.6%) and discipline-specific (46.6%) journals. The number of authors per article was most often fewer than 5 (73.6%), and the number of collaborating institutions was predominantly fewer than 3 (82.5%). Incidences of coauthorship and

collaboration with overseas researchers were few but showed a significant increasing trend in the last decade (**Figure 30**, B). The top 5 conditions studied were general and unspecified (37.9%); endocrine, metabolic, and nutritional (15.2%); circulatory (7.8%); psychological (5.9%); and respiratory (4.9%).

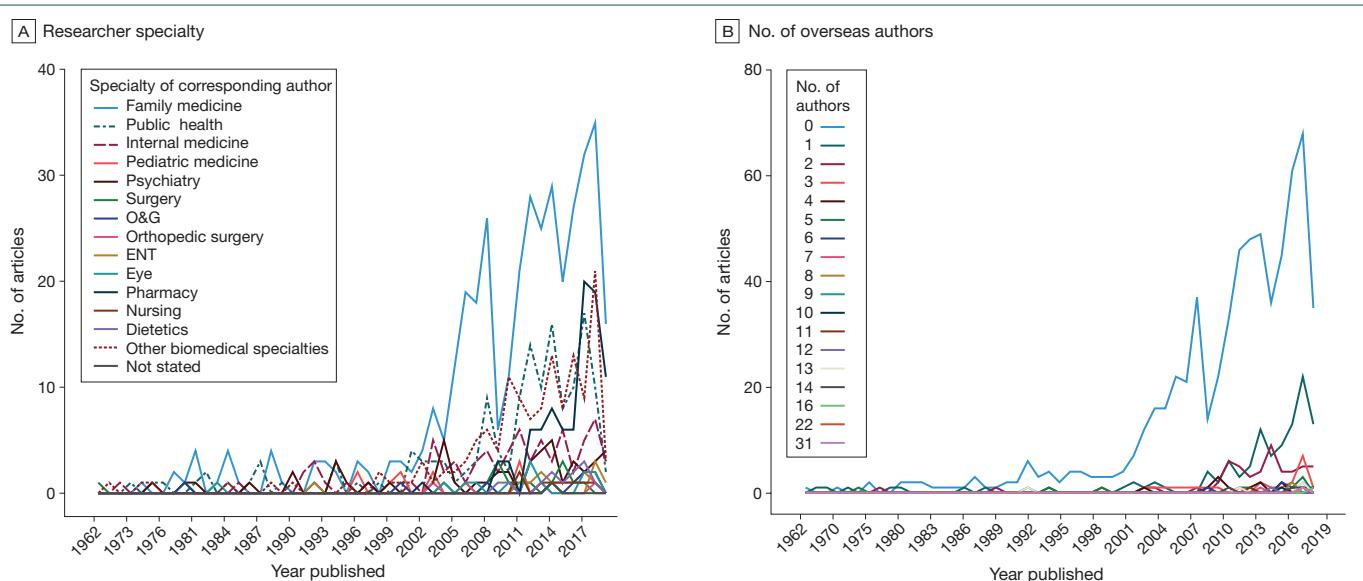
**Conclusions** The longitudinal and prospective trends of the research characteristics assessed in this analysis provided suggestions of improvement initiatives needed for primary care research enterprise in Malaysia. This includes training on the proper use of different study designs, on developing a better supportive ecosystem for interventional clinical trials, on skills for international research collaboration, and for strategizing research topics that meet the issues of primary medical care. Similar works in other disciplines could be initiated and better conducted after this first experience. The aim of phase 2 will be to validate a research-quality screening tool based on domains of relevance, credibility of the methods, and usefulness of the results.

## References

- Chew BH, Lim PY, Lee SWH, et al. A systematic review of medical and clinical research landscapes and quality in Malaysia and Indonesia [REALQUAMI]: the review protocol. PROSPERO identifier: CRD42020152907. Updated April 28, 2020. Accessed July 14, 2022. [https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=152907](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=152907)
- Chew BH, Lim PY, Lee SWH, et al. A systematic review of medical and clinical research landscapes and quality in Malaysia and Indonesia [REALQUAMI]. OSF Registry for Research on the Responsible Conduct of Research. May 1, 2019. Accessed July 14, 2022. <https://osf.io/w85ce>

<sup>1</sup>Department of Family Medicine, Faculty of Medicine and Health Science, Universiti Putra Malaysia, Selangor, Malaysia, chewboonhow@upm.edu.my; <sup>2</sup>School of Pharmacy, Monash University Malaysia, Jalan Lagoon Selatan, Selangor, Malaysia;

**Figure 30. Characteristics of Clinical and Biomedical Research Performed in Malaysia by Malaysian Authors, 1962 to 2017**



ENT indicates ear, nose, and throat; O&G, obstetrics and gynecology.

<sup>3</sup>Department of Community Health, Faculty of Medicine and Health Science, Universiti Putra Malaysia, Selangor, Malaysia; <sup>4</sup>Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia, Penang, Malaysia; <sup>5</sup>Department of Psychiatry, Faculty of Medicine and Health Science, Universiti Putra Malaysia, Selangor, Malaysia; <sup>6</sup>Department of Biomedical Sciences, Faculty of Medicine and Health Science, Universiti Putra Malaysia, Selangor, Malaysia; <sup>7</sup>Medical and Health Sciences Library, Faculty of Medicine and Health Science, Universiti Putra Malaysia, Selangor, Malaysia; <sup>8</sup>Department of Community Medicine, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia

**Conflict of Interest Disclosures** None reported.

## Conflict of Interest

### Conflicts of Interest in Systematic Reviews on Methylphenidate for Attention-Deficit Disorder

Alexandra Snellman,<sup>1</sup> Stella Carlberg,<sup>1</sup> Louise Olsson<sup>1</sup>

**Objective** To compare financial conflict of interest (COI) as declared by authors of systematic reviews (SRs) of methylphenidate for attention-deficit disorder with publicly available information and in relation to risk of bias (RoB).

**Design** SRs on the outcomes associated with methylphenidate for attention-deficit disorder in all ages were searched in Medline, Cochrane, Embase, and PsycInfo. The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guideline was followed for the selection of relevant SRs. Two reviewers (A.S. and S.C.) independently screened open websites and recent publications for all authors of each SR, and data on financial COI were extracted. All searches followed a preplanned and similar routine. A time limit of 3 years from publication of the index SR was applied. Findings were discussed between the reviewers and repeated until consensus was reached. If no data were found for any of the authors, the SR was categorized as no COI. Two reviewers (A.S. and L.O.) independently judged RoB of the SR using the Risk of Bias in Systematic Reviews (ROBIS) tool. Any disagreement was resolved in consensus. In addition, data were retrieved on COI as declared by authors.

**Results** Of 651 unique publications, 44 relevant SRs published between 2008 and 2021 were included. In all, 32 SRs (73%) were based on randomized clinical trials only, 18 (41%) reported positive effects only, 15 (34%) reported both positive and adverse effects, and 11 (25%) reported adverse effects only. Eleven SRs (25%) included only studies using placebo for comparison. A meta-analysis was conducted in 26 of the SRs (59%). COI disclosure was missing for 2 of 44 SRs (5%). For 15 SRs (34%), authors declared COI, and this declaration was confirmed by open sources in all cases. For 27 SRs (61%), the authors declared no COI, but discordant information was publicly available for 8 of 27 (30%). The direction of COI for most SRs was not able to be assessed. The RoB was high in 37 of 44 SRs (84%). Of the 7 SRs with low RoB, 1 had no COI identified in open sources (**Table 78**).

**Table 78. COI Declared by Authors and Open Sources vs RoB in 44 SRs**

	High RoB (n = 35)	Low RoB (n = 7)	Total (N = 42) <sup>a</sup>
<b>COI declared by authors</b>			
Yes	11	5	16
No	24	2	26
<b>COI declared by open sources</b>			
Yes	18	6	24
No	17	1	18

Abbreviations: COI, conflict of interest; RoB, risk of bias; SRs, systematic reviews.

<sup>a</sup>Two SRs lacked disclosures.

**Conclusions** The findings indicated an underreporting of COI in SRs in studies on attention-deficit disorder, and most SRs were compromised by high RoB. Owing to small numbers, no firm conclusion on the association between COI and RoB was possible.

<sup>1</sup>Centre for Assessment of Medical Technology in Örebro, Örebro University Hospital, Örebro, Sweden, alexandra.snellman@regionorebrolan.se

**Conflict of Interest Disclosures** None reported.

## Diversity, Equity, and Inclusion

### Qualitative Assessment of an Antiracism Editorial Internship Program for Early Career Underrepresented Scholars at *Teaching and Learning in Medicine*

Tasha R. Wyatt,<sup>1</sup> Justin L. Bullock,<sup>2</sup> Anna T. Cianciolo,<sup>3</sup> Gareth Gingell,<sup>4</sup> Anabelle Andon,<sup>5</sup> Heeyoung Han,<sup>3</sup> Carlos Torres,<sup>6</sup> Erica J. Odukoya,<sup>7</sup> Elza Mylona,<sup>8</sup> Dario Torre,<sup>9</sup> Zareen Zaidi<sup>10</sup>

**Objective** As gatekeepers for knowledge production, editors use their backgrounds to determine what counts as high-quality research, how research should be done, and whether it offers novel insights. Individuals who come from racialized backgrounds are often excluded from these conversations, ensuring the publishing world continues to be led by the perspectives of the dominant racial group.<sup>1</sup> To address this issue, the journal *Teaching and Learning in Medicine* (TLM) created an internship program for early career scholars who self-identified as Black, Latinx, or Indigenous to gain experience in the publishing/editorial process.<sup>2</sup> The purpose of this study was to critically examine the outcomes of the first 6 months and investigate its implications for other journals interested in similar efforts.

**Design** A collaborative autoethnographical study was designed to collectively analyze the program. Editorial staff and interns reflected on their lived experiences participating in the internship to understand the current culture of academic publishing and explore how to improve.<sup>3</sup> Data sources included archival emails and program planning documents, focus group data, and group exercises. After transcribing the focus groups, data were analyzed using

thematic analysis by both editorial staff and interns. Institutional review board approval was not required for this study.

**Results** Focus group data indicated that while *TLM* designed the program with antiracist intent, mentors did not foreground their race in the editorial/publishing process outside of the intern's minoritized status as a program selection criterion. Early on, mentors viewed the internship more as an opportunity to improve the journal's editorial work, for participants to gain experience working in a historically guarded space, and to contribute to achieving racial equity in medical education. Therefore, despite the program's specific antiracist focus, 6 months in, none of the mentors (who identified as White individuals) had explicitly discussed topics of race with their interns. Mentors' reasons for not discussing race varied, including uncertainty about how to invite interns into such a discussion and not seeing interns as racialized individuals. However, at the end of the 6 months, researchers realized the need to discuss this topic, thus moving the program into explicit conversations about race and the role it plays in publishing.

**Conclusions** Although the program met some antiracist goals, stakeholders did not explicitly discuss the role that race plays in the review process, therefore limiting the program's initial impact. Through this collaborative autoethnography, *TLM* stakeholders critically reflected on the program in real time and addressed this gap. In doing so, they engaged in the ongoing critical action needed to support equity within the editorial process. Editorial staff at *TLM* now have the opportunity to address ongoing power dynamics between interns and staff members that will advance the journal's efforts at antiracism. Plans are currently underway to ensure interns' experiences are incorporated.

## References

1. Salazar JW, Claytor JD, Habib AR, Guduguntla V, Redberg RF. Gender, race, ethnicity, and sexual orientation of editors at leading medical and scientific journals: a cross-sectional survey. *JAMA Intern Med.* 2021;181(9):1248-1251. doi:10.1001/jamainternmed.2021.2363
2. Cianciolo AT. Letter from the editor—*Teaching & Learning in Medicine*'s anti-racism strategy. *Teach Learn Med.* 2020;32(5):457-458. doi:10.1080/10401334.2020.1836461
3. Chang H, Ngunjiri FW, Hernandez KAC. *Collaborative Autoethnography*. Left Coast Press; 2013.

<sup>1</sup>Center for Health Professions Education, Uniformed Services University, Bethesda, MD, USA, tasha.wyatt@usuhs.edu;  
<sup>2</sup>University of California, San Francisco, San Francisco, CA, USA;  
<sup>3</sup>Southern Illinois University School of Medicine, Springfield, IL, USA; <sup>4</sup>The University of Texas at Austin, Austin, TX, USA; <sup>5</sup>City University of New York School of Medicine, New York, NY, USA;  
<sup>6</sup>Harvard University, Boston, MA, USA; <sup>7</sup>University of Michigan, Ann Arbor, MI, USA; <sup>8</sup>Eastern Virginia Medical School, Norfolk, VA, USA; <sup>9</sup>University of Central Florida College of Medicine, Orlando, FL, USA; <sup>10</sup>George Washington University, Washington, DC, USA

**Conflict of Interest Disclosures** None reported.

**Disclaimer** This work was prepared by a military or civilian employee of the US government as part of the individual's official duties and therefore is in the public domain. The opinions and assertions expressed herein are those of the author(s) and do not necessarily reflect the official policy or position of the Uniformed Services University or the US Department of Defense.

**Additional Information** Anabelle Andon is a co-corresponding author.

## Editorial and Peer Review Process

### Analysis of Timing of Manuscript Submissions and Assignment of Editors and Reviewers on Editorial Decisions at *eLife*

Weixin Liang,<sup>1</sup> Kyle Mahowald,<sup>2</sup> Jennifer Raymond,<sup>3</sup> Vamshi Krishna,<sup>4</sup> Daniel Smith,<sup>4</sup> Dan Jurafsky,<sup>1,5</sup> Daniel McFarland,<sup>4</sup> James Zou<sup>1,6,7</sup>

**Objective** Editorial decisions can depend on factors, like the timing of submissions or the matching of editors and reviewers, that are independent of the quality of the work. This analysis investigates associations of these and other external factors with editorial outcomes at *eLife*, a major biomedical journal.

**Design** This study analyzed whether timing of submission (weekend vs weekday) was associated with the decision to send manuscripts for external review among senior editors, who can desk-reject submissions; compared peer review manuscript ratings by reviewer volume; and assessed whether submission time of month or year, preceding decisions to reject or review a manuscript, and matching of reviewer to manuscript specialty influenced editorial decisions. Data were analyzed by single variable regression.

**Results** Between January 2016 and December 2018 *eLife* received 23,190 total submissions, 6498 of which were sent for review. Among senior editors ( $n = 65$ ), proportions of manuscripts sent for external review ranged from 9.6% to 49.3% and were statistically significantly lower on weekends (mean, 24% [SD, 1.3%]) than on weekdays (mean, 29% [SD, 0.6%];  $P < .001$ ), an association observed for most senior editors. Average peer reviewer rating (range, 0-1) increased with volume category: mean of 0.453 (SD, 0.003) for 1 to 5 submission reviews; 0.463 (SD, 0.008) for 6 to 10 submissions; and 0.472 (SD, 0.007) for 11 or more submissions, and reviewers' ratings increased with successive reviews. In a nonquantitative inspection of submission and decision trends, submission time of month or year, preceding decisions to reject or review a manuscript, and matching of reviewer to manuscript specialty did not appear to influence editorial decisions.

**Conclusions** This study found a statistically significant association between timing of submission during the week and editorial decisions. Peer reviewer ratings increased with review volume. Submission time of month or year, preceding decisions to reject or review a manuscript, and matching of

reviewer to manuscript specialty did not appear to influence editorial decisions.

<sup>1</sup>Department of Computer Science, Stanford University, Stanford, CA, USA, wxliang@stanford.edu; <sup>2</sup>Department of Linguistics, University of California, Santa Barbara, Santa Barbara, CA, USA; <sup>3</sup>Department of Neurobiology, Stanford University, Stanford, CA, USA; <sup>4</sup>Graduate School of Education, Stanford University, Stanford, CA, USA; <sup>5</sup>Department of Linguistics, Stanford University, Stanford, CA, USA; <sup>6</sup>Department of Biomedical Data Science, Stanford University, Stanford, CA, USA; <sup>7</sup>Chan-Zuckerberg Biohub, San Francisco, CA, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This research is supported by National Science Foundation SCISIPBIO (A Science of Science Policy Approach to Analyzing and Innovating the Biomedical Research Enterprise) project 2022435.

**Additional Information** We thank Andy Collings and James Gilbert from *eLife* for sharing the data and providing explanations and *eLife*'s editorial leadership team for input and suggestions.

### Association Between Number of External Peer Review Invites, Unsuccessful Invites, and Declined Reviews With Rejection of Manuscripts

Gene Y-K Ong,<sup>1,2,3</sup> Ellen Weber,<sup>3,4</sup> Joshua McAlpine<sup>5</sup>

**Objective** Many journals encounter difficulties obtaining peer reviewers. However, it is unknown if this is associated with publication outcomes.<sup>1,2</sup> The purpose of this research was to evaluate whether, for initial manuscript submissions that were eventually peer reviewed, there was a significant association between the number of external peer reviewers sought, unsuccessful invites, and declined invites for external peer review and the publication outcome.

**Design** This was a retrospective study of anonymized, unique, original research submissions to the *Emergency Medicine Journal* (*EMJ*) that received at least 1 external peer review over a 5-year period (January 2016 to December 2020). A database of deidentified original manuscripts submitted to *EMJ* during the study period was interrogated to

determine if there were significant associations between the number of unsuccessful external peer review invitations, the number of total invitations needed, and the acceptance or rejection of a research manuscript. Original submissions without any external peer review were excluded. Statistical review invitations were excluded from the data. Mann-Whitney *U* test was used to assess differences between the variables and their publication outcomes. Odds ratios (OR), likelihood ratios (LR), and positive predictive value (PPV) were used as measures of association for potential thresholds for variables and publication outcomes.

**Results** There were 806 deidentified peer-reviewed original submissions included with 85 manuscripts (10.6%) accepted for publication during the study period (Table 79). The ORs for a peer-reviewed original research submission eventually being rejected, according to number of invitations, were 77.5 (95% CI, 24.2-248.2; *P* < .001; LR, 21.0; 95% CI, 6.9-63.7; PPV, 99.4%; 95% CI, 98.3-99.8) for submissions with 4 or more invitations, 36.8 (95% CI, 9.0-150.9; *P* < .001; LR, 20.0; 95% CI, 5.1-78.8; PPV, 99.4%; 95% CI, 97.7-99.9) for submissions with 3 or more unsuccessful external peer review invitations, and 22.6 (95% CI, 3.1-163.8; LR, 18.0; 95% CI, 2.6-127.2; PPV, 99.4%; 95% CI, 95.6-99.9) for submissions with 2 or more peer reviewers who declined review invitations.

**Conclusions** The number of declined peer review invitations and total review invitations prior to a decision were associated with rejection of a manuscript. The wide 95% CIs in these results could be due to the high variability of underlying factors that could have influenced the difficulty in getting peer reviews and their interplay with the decision to publish. The findings of this study may also be potentially different for different journals.<sup>3</sup> Further research should be done to provide further insights on specific factors that may be associated with difficulties in getting peer reviewers.<sup>1-3</sup>

**Table 79. Characteristics of Peer-Reviewed Original Article Submissions and Their Eventual Publication Outcomes**

Submitted initial manuscripts with $\geq 1$ external peer review (N = 806)	External peer reviewer invites for initial manuscript submission	External peer reviewer invites without a review	Unsuccessful external peer review invites with response				External peer review invites without a response <sup>a</sup>
			Total	Declined	Unavailable	Conflict of interest	
Rejected, total No. (n = 721)	4560	3084	1484	626	633	125	1600
Accepted, total No. (n = 85)	177	58	14	6	7	1	44
Rejected, range	1-27	0-24	0-12	0-7	0-8	0-3	0-19
Accepted, range	1-9	0-8	0-4	0-2	0-2	0-1	0-5
Rejected, mean (SD)	6.32 (4.14)	4.28 (4.05)	2.06 (2.23)	0.87 (1.19)	1.02 (1.28)	0.17 (0.48)	2.22 (2.54)
Accepted, mean (SD)	2.08 (1.25)	0.66 (0.11)	0.16 (0.59)	0.07 (0.30)	0.08 (0.31)	0.01 (0.11)	0.49 (0.82)
Rejected, median (IQR)	5 (3-8)	3 (1-6)	1 (0-3)	0 (0-1)	1 (0-2)	0	2 (0-3)
Accepted, median (IQR)	2 (2-2)	0 (0-1)	0 (0-1)	0	0	0	0 (0-1)
<i>P</i> value <sup>b</sup>	<.001	<.001	<.001	<.001	<.001	<.001	<.001

<sup>a</sup>No response or automatically declined.

<sup>b</sup>2-Tailed, asymmetric *P* values calculated with Mann-Whitney *U* test.

## References

1. Tsvina A, Spelley R, Palatnik A. Bias in the peer review process: can we do better? *Obstet Gynecol.* 2019;133(6):1081-1083. doi:10.1097/AOG.oooooooooooo03260
2. Tennant JP, Dugan JM, Graziotin D, et al. A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Res.* 2017;6:1151. doi:10.12688/f1000research.12037.3
3. Stamm T, Meyer U, Wiesmann HP, Kleinheinz J, Cehreli M, Cehreli ZC. A retrospective analysis of submissions, acceptance rate, open peer review operations, and prepublication bias of the multidisciplinary open access journal. *Head Face Med.* 2007;3:27. doi:10.1186/1746-160X-3-27

<sup>1</sup>Department of Emergency Medicine, KK Women's and Children's Hospital, Singapore, geneong@yahoo.com; <sup>2</sup>Duke-National University of Singapore Medical School, Singapore; <sup>3</sup>*Emergency Medical Journal*, London, UK; <sup>4</sup>Department of Emergency Medicine, University of California, San Francisco, San Francisco, CA, USA; <sup>5</sup>BMJ Publishing Group, London, UK

**Conflict of Interest Disclosures** None reported.

## Authors' General Experiences With Submitting Manuscripts and With Submission Prefill to a Manuscript Submission System for the *Annals of African Surgery*

Vincent Kipkorir,<sup>1</sup> Ernest Kimani,<sup>1</sup> James Kigera<sup>1</sup>

**Objective** Over the past decade, several journals in Africa have transitioned from email submission systems to web-based platforms for manuscript submission and editorial processes.<sup>1</sup> Previous reports<sup>1,2</sup> assessing experience with these systems have revealed conflicting feedback from authors and editorial teams, with varying reports of user friendliness. Notably, a letter to the editor published in *Nature*<sup>3</sup> stated that such systems "make authors do all the work." Over time, online submission systems have sought to have in-built programs to enable automation of the submission process to ease the authors' experience. The *Annals of African Surgery* uses the submission prefill feature in its submission system for the aforementioned purpose.

**Design** This study was cross-sectional. To establish efficacy and gather authors' experience of the online submission system and the submission prefill (ie, the system that automatically extracts and populates submission fields for authors from an uploaded manuscript text file), authors submitting articles to the *Annals of African Surgery* were surveyed between September and December 2021. This survey was delivered to authors using the manuscript submission system, during the submission process, and at the end of the editorial workflow after the final decision was made.

**Results** A total of 44 authors completed the survey on submission prefill, and 42 completed the survey on the manuscript submission system. Because the surveys were

embedded as part of the submission processes, the response rate was 100%. Overall, the responses to the question of whether the submission prefill was effective in reducing the workload were divided evenly (50.0% agree vs 50.0% neutral), but there was overwhelmingly positive feedback on the effectiveness of manuscript submission system in terms of its ease of use (88.1%) and in obtaining the status of the manuscript (92.9%). The survey questions and their respective responses are as outlined in **Table 80**.

**Conclusions** There was overall positive feedback regarding authors' interaction with the web-based online submission interface. Feedback from authors' experience on whether submission prefills reduce submission labor, however, still remains divided. Measures taken to improve the prefill's capacity to extract manuscript details with better accuracy would, hence, enhance authors' experience.

## References

1. Ware M. Online submission and peer-review systems. *Learned Publishing.* 2005;18(4):245-250. doi:10.1087/095315105774648771
2. Ware M. Authors say that they prefer online submission. *Nature.* 2005;434(7033):559. doi:10.1038/434559d
3. Moore JP. Online submission makes authors do all the work. *Nature.* 2005;433(7028):800. doi:10.1038/433800d

<sup>1</sup>The *Annals of African Surgery*, Nairobi, Kenya, vincentkipkorir42357@gmail.com, vincent.kipkorir@annalsofafricansurgery.com

**Conflict of Interest Disclosures** James Kigera is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. No other disclosures were reported.

**Table 80. Survey Questions and Author Responses**

Survey question	Responses, No. (%)		
	Yes	No	Neutral
<b>Submission prefill (n = 44)</b>			
Did you use the submission prefill during your submission?	22 (50.0)	22 (50.0)	NA
	Agree	Disagree	Neutral
The submission prefill extracted the information without errors and with minimal correction needed.	22 (50.0)	NA	22 (50.0)
The submission prefill reduced the work of manually uploading each segment of the manuscript.	22 (50.0)	NA	22 (50.0)
Would you recommend others to use the submission prefill system?	22 (50.0)	NA	22 (50.0)
<b>Manuscript submission system (n = 42)</b>	Agree	Disagree	Neutral
The manuscript submission system is quick and easy to use.	37 (88.1)	3 (7.1)	2 (4.8)
There was sufficient communication about the status of my paper.	39 (92.9)	1 (2.4)	2 (4.8)

Abbreviation: NA, not applicable.

## Assessment of Use of Dedicated Editors for Handling and Reviewing Manuscripts With Previously Obtained Peer Reviews

Riaz Qureshi,<sup>1</sup> Kirsty Loudon,<sup>2</sup> Alexander Gough,<sup>3</sup> Shaun Treweek,<sup>4</sup> Tianjing Li<sup>1</sup>

**Objective** The objective of this research was to assess whether an editorial workflow wherein submissions are handled solely by dedicated protocol editors reduces the average time from initial submission to a final decision compared with standard peer review and whether this reduction differs by protocol formatting (structured or unstructured).

**Design** A cross-sectional study of the workflow timing for protocol submissions to *Trials* was conducted. Type A protocols demonstrate prior peer review as part of their funding process by providing previous comments and self-declaring this prior review in the submission process. Type A protocols are handled by protocol editors as the sole reviewers, whereas type B protocols are handled by other editors and require full peer review (ie, 2 to 3 reviewers). The workflow data stored in the *Trials* manuscript submission and review system were extracted to compare the timing from initial submission to first and final decisions for all protocols, separated by type, before and after protocol editors were implemented (January to December 2019 and January 2020 to November 2021, respectively). The timing was also extracted for submissions that were formatted as unstructured protocols with an accompanying SPIRIT checklist or following the Structured Protocol Template recommended by *Trials*. The workflow timing results are descriptively summarized.

**Results** From January 2020 to November 2021, 1114 type A (360 [32%]) and type B (754 [68%]) trial protocols were submitted to *Trials*. Compared with type B protocols, 143 of which (19%) used a structured template, 137 type A protocols (38%) followed a structured format, possibly owing to

self-selection or prior reviews leading to a more structured approach in submission. A timeline of the workflow for each protocol type by format as well as overall for each period is shown in **Figure 31**. Overall, type A protocols had a mean of 118 days to final decision (from initial submission), whereas type B protocols had a mean of 165 days to final decision. Overall, across protocols in the period after protocol editors were implemented, the time to final decision was reduced 30 days from the preceding year: 146 days (January 2020 to November 2021) versus 176 days (January to December 2019). Among type A protocols, a structured template format reduced the time to final decision compared with unstructured—an association that was not seen in type B protocols, possibly because multiple reviewers were required.

**Conclusions** The use of dedicated editors to handle trial protocols that had already undergone peer review was associated with fewer days to final decision compared with protocols that were handled by other editors and required full peer review. Future work should examine the quality of peer review in both workflows to determine whether use of dedicated reviewers improves the quality of published works as well as the time to publication.

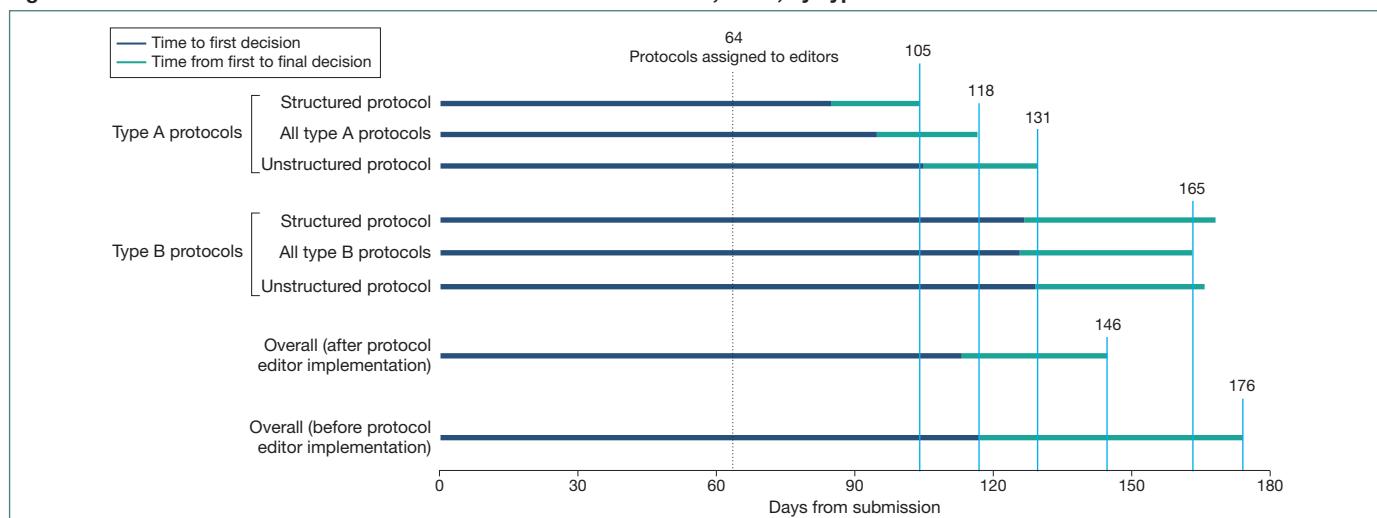
<sup>1</sup>Department of Ophthalmology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; riaz.qureshi@cuanschutz.edu;

<sup>2</sup>Freelance researcher/editor; <sup>3</sup>Institute of Applied Health Research, University of Birmingham, Birmingham, UK; <sup>4</sup>Health Services Research Unit, University of Aberdeen, Aberdeen, UK

**Conflict of Interest Disclosures** Riaz Qureshi, Kirsty Loudon, and Alexander Gough are protocol editors of *Trials*, BMC. Shaun Treweek and Tianjing Li are editors in chief of *Trials*, BMC.

**Acknowledgments** We would like to thank Eleanor Cox, Bex Chang, and Krishna Vairamani for their work as editorial staff and for managing the workflow of peer review for *Trials*.

**Figure 31. Timeline of Decisions for Trial Protocols Submitted to *Trials*, BMC, by Type**



Type A protocols were peer reviewed as part of the funding process (presubmission) and are handled and reviewed by a designated protocol editor. Type B protocols require full peer review and are handled by a nonprotocol editor (eg, associate editor).

## Automatic Classification of Peer Review Recommendation

Diego Kozlowski,<sup>1</sup> Clara Boothby,<sup>2</sup> Rosemary Steup,<sup>2</sup> Pei-Ying Chen,<sup>2</sup> Vincent Larivière,<sup>3,4</sup> Cassidy R. Sugimoto<sup>5</sup>

**Objective** Peer review plays a fundamental role in scholarly publishing, but its legitimacy has been increasingly questioned. A growing literature discusses how reviewers' demographic characteristics and biases might lead to disparities in research dissemination.<sup>1,2</sup> Because the extent to which reviewers are able to determine the outcomes for papers may vary, it is important to look at the relationship between reviewers' recommendations and editors' decision-making. However, reviewer recommendations are often embedded in the text of the review. This work proposes a method for automatic detection of recommendations based on review text.

**Design** The automatic classification used a rule-based algorithm that searched for the presence of 1 or more phrases that signal the reviewer's recommendation: accept, minor revision, major revision, or reject categories, as defined on the hand-coding process. The algorithm considered the different combinations of signal phrases to define the outcome. The list of signal phrases was iteratively built on 3 rounds of hand-coding and fuzzy matching sentences, while the combinations were defined to maximize the precision, on the hand-coded cases. This study used Publons' data set, which contained 3,310,791 reviews from 25,934 journals; while 600 cases were hand-coded, a subset of 200 reviews was used to evaluate the performance. The gender of reviewers was inferred by matching first and last names to curated lists of country-specific gendered names, including the US Census.<sup>3</sup>

**Results** The overall accuracy on the test was that 81% of assigned recommendations were correct according to hand coding ( $n = 149$ ). Since the inclusion of additional phrases is associated with lowered accuracy, this might indicate an upper bound in our experiment, given the limits of the current data and the idiosyncrasies of peer review language. Nonetheless, the algorithm's accuracy was comparable to the rate of agreement between human hand coders ( $n = 60$  [88%]). Over the full data, 14.3% of reviews were assigned a recommendation by this method ( $n = 473,443$ ). This was comparable to the hand-coded identification of 18.3% of reviews containing an explicit recommendation ( $n = 399$ ). From these results, we concluded that the inclusion of an explicit recommendation remains relatively uncommon in peer review, with the majority of peer reviewers leaving a final decision on the manuscript as the responsibility of the editor, but there was large variation between journals. Initial results nonetheless showed gender differences in reviewing behavior, with higher retrieval rates associated with reviewers who identified as men.

**Conclusions** This work is among the first benchmarks for automatic classification of review recommendations on a large-scale, cross-domain database. Though preliminary, it paves the way for future developments, including studies of

potential biases and inequalities in scholarly publishing through examination of the relationship between reviewer characteristics and review outcomes.

## References

1. Lee CJ, Sugimoto CR, Zhang G, Cronin B. Bias in peer review. *J Am Soc Inf Sci Tec.* 2013;64(1):2-17. doi:10.1002/asi.22784
2. Sun M, Barry Danfa J, Teplitskiy M. Does double-blind peer review reduce bias? evidence from a top computer science conference. *J Am Soc Inf Sci Tec.* 2022;73(6):811-819. doi:10.1002/asi.24582
3. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: global gender disparities in science. *Nature.* 2013;504:211-213. doi:10.1038/504211a

<sup>1</sup>Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; <sup>2</sup>Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA, crboothb@iu.edu; <sup>3</sup>École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal, QC, Canada; <sup>4</sup>Observatoire des Sciences et des Technologies, Université du Québec à Montréal, Montreal, QC, Canada; <sup>5</sup>School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported by the Doctoral Training Unit on Data-Driven Computational Modelling and Applications (DRIVEN), which is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781).

**Acknowledgment** We thank Brad Demarest and Chaoqun Ni for contributions to an earlier phase of the project. We also thank the team at Publons who implemented the gender inference algorithm on their data before giving us access.

**Additional Information** Diego Kozlowski is a co-corresponding author.

## Errors and Corrections

### Assessment of Errors in Peer Reviews Published With Articles in *The BMJ*

Fred Arthur<sup>1</sup>

**Objective** It is possible that most published research findings are false.<sup>1</sup> Effective peer review should detect the study error and the mental model error. The study error can be detected by traditional epidemiological/statistical analysis and exists at the level of the study. The mental model error exists at the level of underlying structural intellectual knowledge and beliefs,<sup>2</sup> in which accuracy provides the prior probability and thus substantially affects the posterior probability of the study.

**Design** The peer review process of *The BMJ* was assessed for an inclusive sample of articles published with peer review cycles (ie, editor and reviewer and author comments and decision letters) in the research section in July 2021. Likert

5-point scales were used. The peer-review process for each article was rated for the awareness within the process to detect and manage study error and mental model error, with 5 representing high estimated awareness and efficacy. For mental model error detection, higher Likert scores were accorded in peer review cycles containing the following 4 elements: (1) active criticism of the underlying mental model (possible errors in diagnostic categories, underlying beliefs of therapy mechanisms, and the pathological appropriateness of outcome measures) to estimate the final study posterior probability; (2) awareness that studies produce contingent fact claims and that intellectual validity requires analysis of this contingency; (3) awareness that key patient outcomes must exist outside the mental model to minimize outcome contamination by the model and circular reasoning within the model (eg, use of all-cause morbidity and mortality); and (4) awareness that clinicians are most interested in the probability of the mental model because they use the model daily to inform numerous data decisions made in clinical practice (eg, decisions regarding history, physical examination, differential diagnosis, result, treatment planning, and outcome assessments data). A single ranking was determined for the totality of the review process for each article provided by editors and reviewers and author responses. Result populations of study error and mental model error scores were compared to assess the separation of the distributions using the Mann-Whitney test.

**Results** Forty-five peer reviewers produced 21 rounds of review for 12 articles (**Table 81**). The study error median score was 4.5, and the mental model error median score was 2 with the Mann-Whitney rejection of equal distributions at  $P < .05$ .

**Conclusions** This preliminary study found *The BMJ* peer review process to be effective at detecting and managing study errors but much less aware of mental model error problems. Given that the underlying mental model

**Table 81. Assessment of *The BMJ* Peer Review**

<i>The BMJ</i> article year;volume:elocator	Review rounds, No.	Peer, No.	Study error Likert score	Mental model error Likert score
2021;373:n1421	2	4	5	4
2021;374:n1448	3	5	5	1
2021;373:n1412	2	2	5	1
2021;374:n1506	1	5	4	4
2021;374:n1446	2	4	5	2
2021;374:n1511	3	6	5	5
2021;374:n1537	1	2	3	3
2021;374:n1592	1	4	4	2
2021;374:n1554	2	4	5	1
2021;374:n1585	1	2	4	2
2021;374:n1637	1	4	4	1
2021;374:n1647	2	3	4	3

probability may affect each study's posterior probability, excellent study internal validity may not secure a strong warrant for clinical applicability. Peer review that detects and manages both study error and mental model error may improve clinical study applicability. Preliminary discussions and later guideline formation may help correct this deficiency.

## References

1. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
2. Holland JH, Holyoak KJ, Nisbett RE, Thagard PR. *Induction: Processes of Inference, Learning, and Discovery*. Massachusetts Institute of Technology; 1986.

<sup>1</sup>Western University, London, ON, Canada, fred.arthur@rogers.com

**Conflict of Interest Disclosures** None reported.

## Ethics and Ethical Concerns

### Assessment of Withdrawal of Manuscripts Submitted to the *Journal of Clinical and Diagnostic Research*

Sunanda Das,<sup>1</sup> Aarti Garg,<sup>1</sup> Hemant Jain<sup>1</sup>

**Objective** To study the manuscripts submitted to the *Journal of Clinical and Diagnostic Research (JCDR)* that were withdrawn or remained unattended by the authors after receiving the external peer reviewer's decision, and to further evaluate the authors' characteristics and reasons for withdrawal stated in the correspondence.

**Design** *JCDR* is a broad-specialty, English-language, peer-reviewed, monthly journal with no article processing charges. This cross-sectional study was conducted on manuscripts that were withdrawn or not followed up (ie, abandoned) by authors between January 1, 2018, and December 31, 2021, after posting the external peer review report requesting revision. The manuscripts reported clinical trials, observational studies, analytical studies, case reports, literature reviews, systemic reviews, and meta-analyses. The following data were recorded: dates of submission and the email asking for a withdrawal, designation of the corresponding author, whether a department head or professor was among the authors, and the reason specified in the emails. As a policy at *JCDR*, all the manuscripts that are under process are considered withdrawn if there is no reply from the author for more than 4 months since the last correspondence.

**Results** There were 1080 manuscripts that were withdrawn ( $n = 150$ ) or abandoned ( $n = 930$ ) during the 4-year study period. Among these 1080 manuscripts, only 150 (13.8%) that had an email from the corresponding author requesting a withdrawal were further analyzed in the study. Of these 150 manuscripts, one-third (51 manuscripts [34.0%]) had a professor as the corresponding author, whereas 44 (29.3%)

had a department head or professor among the coauthors. Thirteen manuscripts (8.6%) were withdrawn within a month of the submission despite a quick response from the journal, whereas 79 (52.6%) were withdrawn after more than 3 months with 1 or more revisions. Sixty-five emails (43.3%) did not mention any specific reason for withdrawal, 35 (23.3%) stated that the authors were not willing to or unable to revise the manuscripts as per the reviewers' feedback, and 17 emails (11.3%) stated financial constraints as the reason for withdrawal.

**Conclusions** It is a waste of journal resources and peer reviewers' labor when manuscripts are abandoned or withdrawn. As a mark of good conduct, authors are expected to have forthright communication with journals. If strong reasons exist, authors may withdraw their manuscripts, but clear communication is welcome. Notably, experienced academicians (professors) also share the burden of such misconduct. The effect of such withdrawals is harder for journals that do not charge up-front processing fees. Along with rising inflation, such withdrawals add to the economic burden on a stand-alone journal.

<sup>1</sup>Journal of Clinical and Diagnostic Research, Delhi, India  
(drhemantjain@jcdr.net)

**Conflict of Interest Disclosures** None reported

## Funding/Grant Peer Review

### Assessment of Grant Peer Reviewers' Tolerance for Risk in Research Proposals

Stephen A. Gallo,<sup>1</sup> Karen B. Schmaling<sup>2</sup>

**Objective** Grant peer review relies on scientists' ability to evaluate research proposals' quality. Such judgments are sometimes beyond reviewers' discriminatory power, leading to reliance on subjective biases, including preferences for lower-risk, incremental projects.<sup>1,2</sup> However, reviewers' research risk tolerance during grant peer review has not been well studied.

**Design** In late 2020 to early 2021, participants were recruited through email invitation using the American Institute of Biological Sciences' reviewer databases and the National Institutes of Health's panel rosters. A cross-sectional, prospective study was conducted of peer reviewers' numeric evaluations of mock primary reviewers' comments of a theoretical research proposal in which the level and sources of risks and weaknesses were manipulated. Specifically, there were 4 scenarios, with risks associated with either the investigator (limited experience in leading independent research; n = 199), the approach (lacking pilot studies; n = 205), both investigator and approach (n = 201), or neither (control; n = 605). Each participant evaluated the control and 1 randomly chosen manipulated risk scenario (the order of presentation was randomized). Risk tolerance was measured by the 13-item Openness to Experience scale from the NEO Five-Factor Inventory 3.<sup>3</sup>

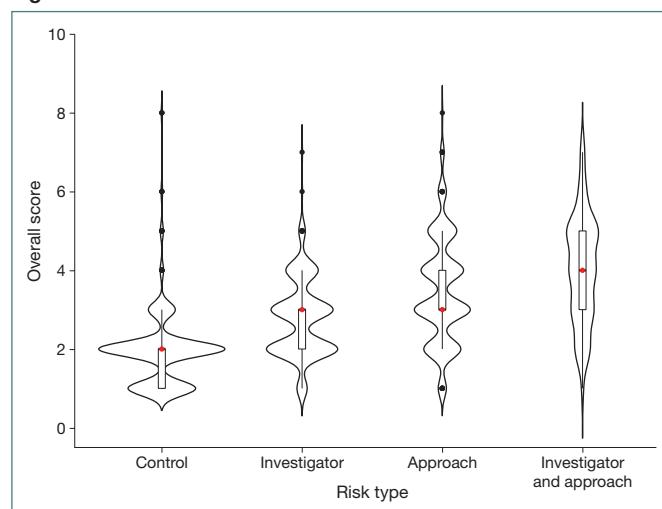
**Results** Violin plots of overall and criteria scoring data from 605 participants (lower scores are more positive) showed that the experimental manipulation was effective: Scenarios with manipulated risk received more mediocre overall scores and more dispersed scoring distributions than the control scenario (**Figure 32**). Ordinal logistic regression models (Nagelkerke  $R^2 = 0.78$ ) suggest that the manipulated risks in these scenarios (compared with control) were strongly associated with participants' overall scores, with odds ratios of 15.0 (95% CI, 3.4-66.0) for the investigator risk scenario, 21.7 (95% CI, 5.5-85.4) for the approach risk scenario, and 36.2 (95% CI, 8.5-154.4) for the scenario with both risks. The approach criterion score was associated with the overall score for all scenarios; significance and innovation criteria scores had stronger associations with overall scores in the control scenario. Openness to experience was not associated with scores, but differing levels of reviewer-reported scoring leniency were observed.

**Conclusions** These data suggest that the evaluation of risks dominates reviewers' evaluation of research proposals and is an important source of interreviewer variability. A weakness is that this study did not involve actual grant proposals. Training peer reviewers to consider risks as potential assets (eg, innovation) may represent an area of useful intervention for science advancement.

### References

1. Lamont M. *How Professors Think: Inside the Curious World of Academic Judgment*. Harvard University Press; 2009.
2. Luukkonen T. Conservatism and risk-taking in peer review: emerging ERC practices. *Res Eval*. 2012;21(1):48-60. doi:10.1093/reseval/rvs001
3. McCrae RR, Costa PT. *NEO Inventories: Professional Manual*. PAR, Inc; 2010.

**Figure 32. Violin Plot**



Probability densities, distributions (via overlaid box plots), and median values (red dots) of the overall scores for different risk scenarios. Boxes indicate IQRs, the whiskers are 1.5 times the IQR plus or minus Q3 or Q1. Black dots represent data points outside the whisker range.

<sup>1</sup>Scientific Peer Advisory and Review Services Division, American Institute of Biological Sciences, Herndon, VA, USA, sgallo@aims.org; <sup>2</sup>Department of Psychology, Washington State University, Vancouver, WA, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported by the Social, Behavioral, and Economic Sciences Office of Multidisciplinary Activities at the National Science Foundation under the standard grant award numbers 1951132 and 1951251.

## Metadata

### Assessing PubMed Metatag Usage for Plain Language Summary Discoverability

Adeline Rosenberg,<sup>1</sup> Slávka Baróniková,<sup>2</sup> William Gattrell,<sup>3</sup> Namit Ghildyal,<sup>4</sup> Tim Koder,<sup>1</sup> Taija Koskenkorva,<sup>5</sup> Andrew Liew,<sup>6</sup> Radha Narayan,<sup>7</sup> Joana Osório,<sup>1</sup> Valérie Philippon,<sup>8</sup> Melissa Shane,<sup>1</sup> Catherine Skobe,<sup>9</sup> Kim Wager<sup>1</sup>

**Objective** PubMed is a popular platform for accessing biomedical research.<sup>1</sup> When tagged correctly, text-based and concise plain language summaries (PLSs) hosted on PubMed can maximize discoverability by a broader audience. This function was introduced in 2019 and allows retrospective tagging of pre-2019 records. Open access (OA) publishing also enhances discoverability, which increases publication accessibility and usage.<sup>2</sup> The aim was to (1) determine the proportion of PubMed records correctly using the PLS tag and reasons for incorrect usage and (2) establish the OA status of journals publishing PLSs on PubMed.

**Design** The entire PubMed database was downloaded (February 9, 2022) and searched for PLSs indexed with an XML <plain-language-summary> tag in the Other Abstract field. Records were deduplicated, and incorrectly tagged PLSs were programmatically excluded for incorrect tag usage (ie, non-PLS content) and confirmed with manual spot checks. Remaining PLSs were categorized by journal and assessed for overall OA status using Journal Selector (Sylogent LLC) or information on journal websites for those PLSs not indexed on Journal Selector.

**Results** There were 3217 records identified with an XML <plain-language-summary> tag in the Other Abstract field, of which just over half (1644 [51.10%]) were published in 2021 (annual prevalence of 0.09% [1644 of 1,769,389]). Of the 3217 records, there were 470 (14.61%) with incorrect tag usage. Categories of incorrect usage included non-English scientific abstract (137 records [4.26%]); duplication of or a greater than 90% similarity score with the scientific abstract in the Abstract field (32 [0.99%]); absence of a scientific abstract in the Abstract field (99 [3.08%]); other non-PLS content (including URLs, novelty statements, and article highlights) (197 [6.12%]); and no content (5 [0.16%]). In addition to these 470 excluded records, there were 124 records using the <plain-language-summary> tag to index both non-English scientific abstracts and English PLSs. Of 105 journals

correctly using the <plain-language-summary> tag for PLSs, 30 (28.57%) were full or gold OA journals, 75 journals (71.43%) offered OA options, and none were closed or subscription-only journals.

**Conclusions** Despite the use of the <plain-language-summary> tag increasing over time,<sup>3</sup> records using this tag represent a minority of all PubMed records, and the tag is used incorrectly for several reasons. There is an unmet need for explicit guidance on both the processes of indexing and the correct usage of the <plain-language-summary> tag, which could help improve correct tagging and uptake. Of note, all PLSs published on PubMed to date come from journals allowing for OA publishing, which aids discoverability and accessibility. Limitations of this analysis include lack of PLS quality assessment, small sample size largely owing to low publisher uptake and tagging, and confounding effects of OA on discoverability. Ultimately, these findings highlight an opportunity for journals to increase the impact of their content and reach a broader audience.

### References

1. Williamson PO, Minter CIJ. Exploring PubMed as a reliable resource for scholarly communications services. *J Med Libr Assoc.* 2019;107(1):16-29. doi:10.5195/jmla.2019.433
  2. Li H, Liu L, Wang X. The open access effect in social media exposure of scholarly articles: a matched-pair analysis. *J Informetrics.* 2021;15(3):101154. doi:10.1016/j.joi.2021.101154
  3. Gattrell W, Wager K, Sheikh N, Chisholm A. Prevalence and characteristics of plain language summaries indexed in PubMed: original abstracts from the 2022 European Meeting of ISMPP. *Curr Med Res Opin.* 2022;38(suppl 1):25-45.
- <sup>1</sup>Oxford PharmaGenesis Ltd, Oxford, UK, adeline.rosenberg@pharmagenesis.com; <sup>2</sup>Galápagos NV, Mechelen, Belgium; <sup>3</sup>Ipsen, Milton Park, UK; <sup>4</sup>Janssen Global Services LLC, Raritan, NJ, USA; <sup>5</sup>Novartis Pharma AG, Basel, Switzerland; <sup>6</sup>Oxford PharmaGenesis Pty Ltd, Melbourne, Australia; <sup>7</sup>Alexion, AstraZeneca Rare Diseases, Boston, MA, USA; <sup>8</sup>Takeda Development Center Americas Inc, Cambridge, MA, USA; <sup>9</sup>Pfizer Inc, New York, NY, USA

**Conflict of Interest Disclosures** At the time of abstract development, William Gattrell was an employee and shareholder of or held stock or stock options in Ipsen, Milton Park, UK, and is now an employee of Bristol Myers Squibb, Uxbridge, UK. Valérie Philippon was an employee and shareholder of or held stock or stock options in Takeda Development Center Americas Inc, Cambridge, MA, USA, and is now an employee of UCB, Cambridge, MA, USA. All other authors are employees of their respective affiliations. Slávka Baróniková, Namit Ghildyal, Radha Narayan, and Catherine Skobe may be shareholders or hold stock or stock options for their respective affiliations.

**Funding/Support** This work was produced by representatives of Open Pharma. Open Pharma is a multisponsor collaboration facilitated by Oxford PharmaGenesis Ltd and received sponsorship funding from Alexion Pharmaceuticals Inc, AstraZeneca Pharmaceuticals LP, Boehringer Ingelheim International GmbH, ER Squibb & Sons LLC, F. Hoffmann-La Roche AG, Galápagos NV, Gilead Sciences Inc, GlaxoSmithKline Biologicals SA, Ipsen Biopharm Ltd, Janssen Global Services LLC, John Wiley & Sons

Ltd, Novartis Pharma AG, Novo Nordisk A/S, Pfizer Inc, Takeda Development Center Americas Inc, and UCB Biopharma SRL.

**Role of the Funder/Sponsor** This abstract represents the work of the individual authors and not of their respective affiliations. The organizations providing sponsorship funding to Open Pharma were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; or decision to submit the abstract for presentation.

**Acknowledgment** Editorial support was provided by Velissaria Vanna of Oxford PharmaGenesis, Oxford, UK.

## Misconduct

### Trends in Research on Plagiarism Among Brazilian Graduate-Level Studies

Renan Almeida<sup>1</sup>

**Objective** Plagiarism is a main concern regarding misconduct in scientific publication. Therefore, it is interesting to gauge and characterize research interests on the topic, particularly in countries with a less established research tradition. The present study updates information on plagiarism research directions among Brazilian graduate-level researchers.

**Design** Two databases were used for this study. The first (CAPES database) is maintained by the country's Ministry of Education, registering information on all master's/doctoral studies in the country.<sup>1</sup> The second (IBICT/BDTD database), a similar database, is maintained by the country's Ministry of Science.<sup>2</sup> Approximately 200,000 manuscripts are included in these databases. The following keywords were used to first identify those manuscripts concerning plagiarism: *plagiarism, science integrity, ethics, and scientific misconduct* (in Portuguese and in English). After manuscripts were identified, abstracts (or the full manuscripts, when available) were read to select those studies specifically dealing with academic plagiarism. Thus, studies on the arts (eg, theater, music, painting), advertisements, en passant mentions, and historical commentaries were not included in the analysis. Items were then classified as the following<sup>3</sup>: conceptual studies (eg, discussions about the concept or the implications of plagiarism), development of detection software and methods, teaching/implementation of prevention methods, legal aspects, and case or quantitative studies.

**Results** A total of 206 theses/dissertations were identified, the first of which was from 1993. Of these, 119 were selected according to the criteria above, and 1 was discarded because it could not be located. Studies pertained to all main areas of academia, with a large number from engineering/computer sciences (eg, "detection"). Numbers for the past 15 years were as follows: 2006 to 2010: 18; 2011 to 2015: 38; 2016 to 2020: 56. The most common research area was development of detection software, with 40 of the 118 identified manuscripts. Conceptual studies followed second (30 of 118), and

prevention methods comprised 18 studies. This tendency was still evident in the last analyzed 5 years, when detection comprised 18 of 56 studies.

**Conclusions** Interest in plagiarism research seems to be increasing in Brazil, a trend that is compatible with the growing concern of its funding agencies with the misconduct topic. In the past decades, the main focus moved away from the discussion of historical events to more applied areas of research. The development of detection software was the most active of these areas, a trend that has been noticed before<sup>3</sup> and continues to this day.

### References

1. CAPES. CAPES Dissertations Repository (in Portuguese). CAPES/Ministry of Education. Accessed January 2022. <https://catalogodetes.capes.gov.br>
2. IBCIT/BDTD. Brazilian Dissertations Library (in Portuguese). IBCIT/BDTD/Ministry of Science. Accessed January 2022. <https://bdtd.ibict.br>
3. Almeida MVR, Fontes-Pereira J. Graduate research on plagiarism in Brazil: characterization of current status. In: *5th World Congress on Research Integrity Abstract Book*. World Conferences on Research Integrity; 2017:94. Accessed January 2022. <https://wcri.org/documents/41-abstract-book-5th-wcri-2017>

<sup>1</sup>Programa de Engenharia Biomédica (PEB)/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, renan@peb.ufrj.br

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This study was funded by the Coordination for the Improvement of Higher Education Personnel—CAPES/Ministry of Education, Brazil, whom we thank.

### Searching for Misconduct and Paper Mills in Peer Review Comments

Adam Day<sup>1</sup>

**Objective** The objective was to test and compare various methods to detect text duplication in peer reviews submitted by 2 or more reviewers.

**Design** Peer review fraud is a significant concern.<sup>1,2</sup> A data set of peer review comments submitted to SAGE Publishing was analyzed to search for duplicate text, a possible sign of fake peer review.<sup>3</sup> Peer review comments for each article peer reviewed by 19 SAGE Publishing journals were downloaded from the ScholarOne peer review management system and loaded into a Pandas DataFrame. Journals were chosen based on the availability of data; therefore, the data set should be considered biased. Similar comments were found using a number of search methods, including MinHash Locality Sensitive Hashing (MinHash LSH) for detecting near-duplicate text strings, and Elasticsearch, a scalable graph database combined with RapidFuzz, a fast string-comparison library, for distinguishing similar from dissimilar comments.

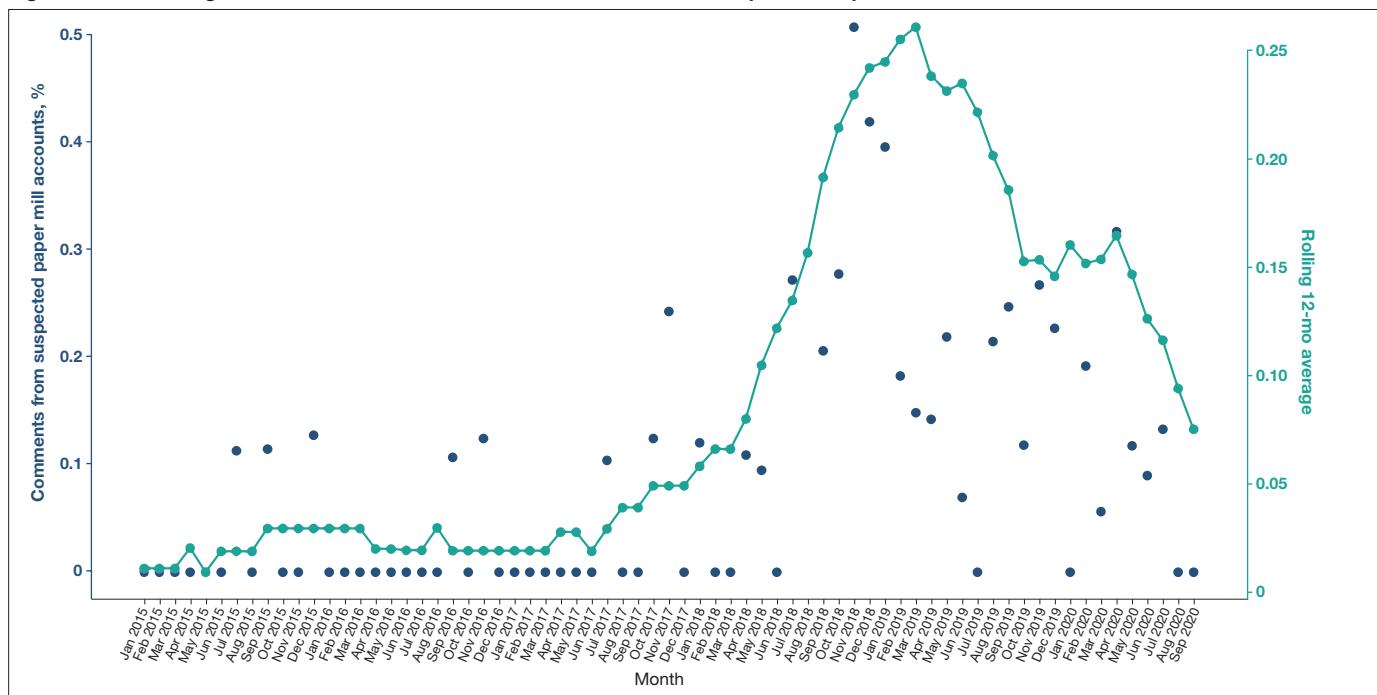
**Results** Of 62,974 peer reviewer accounts used to evaluate 66,815 articles, 357 accounts (0.05%) were identified that produced reviews with partial or fully duplicate comments. One large cluster of 47 accounts that shared a number of reports included a number of articles rejected because of suspected paper mill activity. This number suggests that the cluster of 47 accounts represented 47 fake reviewer accounts administered by a paper mill. In total, 972 articles (1.5%) had reviews from reviewer accounts associated with duplicate commenting activity, and 77 articles had reviews from the 47 suspected paper mill accounts (**Figure 33**). Different search methods identified different suspect accounts and clusters. These searches included (1) a search for exact duplicates, which took 16 seconds to load data into memory and less than 1 second to execute; this search found 29 accounts that had produced similar comments, and (2) a search for similar comments using Elasticsearch, which took 18 minutes and 29 seconds to index and 9 hours, 19 minutes, and 2 seconds to execute; this search found 204 accounts that had produced similar comments.

**Conclusions** Efficient methods for identifying possible peer review fraud and paper mill activity were described. The methods should be tested on broader peer review sets and settings. When duplication is found, the findings must be considered in context before a judgment can be made about whether there is misconduct.

## References

1. Misra DP, Ravindran V, Agarwal V. Integrity of authorship and peer review practices: challenges and opportunities for improvement. *J Korean Med Sci*. 2018;33(46):e287. doi:10.3346/jkms.2018.33.e287

**Figure 33. Percentage of Comments in SAGE Dataset Received From Suspected Paper Mill**



The number of reviews received from suspected paper mill accounts appeared to increase in 2018. There was an apparent decrease in suspected paper mill reports since early 2019. However, this decrease may be associated with a data-quality issue: it is possible that the paper mills have begun to make their fake comments more unique or that some recent template comments have been seen only once and were therefore not identified as duplicates.

2. Cohen A, Pattanaik S, Kumar P, et al. Organised crime against the academic peer review system. *Br J Clin Pharmacol*. 2016;81(6):1012-1017. doi:10.1111/bcpt.12992

3. Dadkhah M, Kahani M, Borchardt G. A method for improving the integrity of peer review. *Sci Eng Ethics*. 2018;24(5):1603-1610. doi:10.1007/s11948-017-9960-9

<sup>1</sup>SAGE Publishing, London, UK

**Conflict of Interest Disclosures** None reported.

## Detection of Plagiarism Using a Search Engine

Ariella Reynolds,<sup>1</sup> Alison Abritis,<sup>2,3</sup> Ivan Oransky<sup>2,4,5</sup>

**Objective** Google has been shown to aid in the detection of plagiarism in text searches<sup>1,2</sup> and in image searches.<sup>3</sup> This pilot study examined the efficacy of using Google to detect matching sources through text, image, and data searches through time trials.

**Design** Data gathering occurred in January and May 2022. Articles retracted for plagiarism (PFound) were pulled from the Retraction Watch Database (<http://retractiondatabase.org>) and listed in an Excel worksheet with distinct numeric identifiers. An Excel random number generator was used to select 17 individual articles. Text, images, and data (as available) from each PFound were tested in locating a potential plagiarism source (PSource). No more than 6 searches were performed for each type of search; searching was stopped once a PSOURCE was detected. Text and data were entered into the Google search box in quotation marks. For text searches, text (generally from the introduction, discussion, or conclusion) was chosen from phrases using

uncommon or unusual word choices, or those exhibiting different language patterns from other text. Strings of data from the results or discussion section were used for data searches. For image searches, images were captured and saved to the computer drive using a screen capture tool; the image was then uploaded to Google image search. To be considered as a PSource, matching results were checked to confirm that PFound had no suitable citation of the PSource, and that PSource was published prior to PFound.

**Results** See **Table 82** for detailed results. Text searches: Text matches were found for all 17 articles; 9 articles required 1 or 2 attempts and only 1 required 6 attempts. Search times ranged from 54 seconds (1 attempt) to 871 seconds (6 attempts); median (IQR) time for all searches was 208 (138.5–289.5) seconds. Successful single search attempts had an average of 9.11 words in the search phrase (excluding stop words); the average fairly steadily decreased per attempt to 3 words. Image searches: Of the 2 articles with image matches, one match was made in the first try (194 seconds) and the other required 4 tries (534 seconds). Mean (SD) time for all searches was 285.54 (98.46) seconds. Data searches: Five articles had data matched in the Google search (mean [SD] search time, 275.8 [100.01] seconds). Three articles required 3 attempts (mean [SD] time, 214.33 [23.18] seconds) and 2 articles required 4 search attempts (mean [SD] time, 368.00 [103.24] seconds)

**Conclusions** Sources of known plagiarism were detected within a mean of 5 minutes using Google. Source material was found for all sample articles using text matching; data string matching occurred more often than image matching.

**Table 82. Summary of Results**

Retracted article	Date published	No. of unique source articles	No. of tries via text	Total time of text search, s	No. of tries via image	Total time of image search, s	No. of tries via data	Total time of data search, s
Article 1	10/28/20	2	2	132	No match	262	No match	164
Article 2	1/12/21	1	3	252	4	534	3	199
Article 3	8/22/17	1	2	145	No match	182	No match	214
Article 4	8/22/21	2	4	442	NA	NA	No match	252
Article 5	7/15/13	1	2	147	NA	NA	No match	319
Article 6	12/10/20	1	2	108	No match	234	No match	274
Article 7	3/15/21	2	1	54	NA	NA	4	295
Article 8	9/20/16	2	4	261	No match	225	3	241
Article 9	4/10/19	1	3	297	No match	311	NA	NA
Article 10	11/24/21	2	2	208	1	194	NA	NA
Article 11	9/30/17	1	4	183	No match	341	No match	347
Article 12	3/2/21	1	6	871	No match	292	No match	699
Article 13	3/17/21	2	3	282	No match	289	4	441
Article 14	8/10/21	2	2	308	No match	299	No match	288
Article 15	7/29/21	1	3	244	NA	NA	3	203
Article 16	5/15/12	1	1	172	No match	164	No match	516
Article 17	9/21/20	1	1	111	No match	385	No match	318
Overall, mean (SD)				248.06 (186.24)		285.54 (98.46)		318 (140.38)

Abbreviation: NA, not applicable.

The number of words used in the search phrase did not appear to influence the search success; the choice of words seemed to be associated with greater matching success.

## References

1. Mondal S, Mondal H. Google search: a simple and free tool to detect plagiarism. *Indian J Vasc Endovasc Surg.* 2018;5:270–273. doi:10.4103/ijves.ijves\_60\_18
2. Holmberg M, McCullough M. Anti-Plagiarism Tools: Scirus v. Google. Purdue University. 2005. Accessed July 11, 2022. <https://docs.lib.purdue.edu/iatul/2005/papers/13>
3. Alawad AA. Exploring Google reverse image search to detect visual plagiarism in interior design. *J High Educ Theory Pract.* 2021;21(10). doi:10.33423/jhetp.v21i10.4634

<sup>1</sup>Avon High School, Avon, CT, USA; <sup>2</sup>Retraction Watch, The Center for Scientific Integrity, New York, NY, USA, abritis@retractionwatch.com; <sup>3</sup>College of Public Health, University of South Florida, Tampa, FL, USA; <sup>4</sup>Arthur Carter Journalism Institute, New York University, New York, NY, USA; <sup>5</sup>Spectrum, The Simons Foundation, New York, NY, USA

**Conflict of Interest Disclosures** Ariella Reynolds and Alison Abritis are a part-time contractor and an employee, respectively, of The Center for Scientific Integrity, which developed and maintains the Retraction Watch Database. Ivan Oransky is the volunteer executive director of The Center for Scientific Integrity, a nonprofit organization that is funded through database licensing fees, a subcontract from the University of Illinois on a Howard Hughes Medical Institute grant, and donations from individuals.

**Additional Information** Ivan Oransky is a co-corresponding author.

## Pandemic Science

### Assessment and Comparison of Preprints and Peer-Reviewed Publications of Reporting Characteristics of Randomized Clinical Trials of Pharmacologic Treatment for COVID-19

Philipp Kapp,<sup>1,2,3,4</sup> Laura Esmail,<sup>1,2,3</sup> Lina Ghosn,<sup>1,2,3</sup> Philippe Ravaud,<sup>1,2,3</sup> Isabelle Boutron<sup>1,2,3</sup>

**Objective** Due to the pandemic, preprint servers contained up to 25% more trials, whereas some medical journals accelerated their editorial processes to ensure the rapid dissemination of findings.<sup>1</sup> Concerns regarding quality and transparency rose.<sup>2</sup> This meta-study assessed the transparency, completeness, and consistency of COVID-19 reports and whether there was an improvement after journal peer review.

**Design** The Cochrane COVID-19 Study Register and L-OVE COVID-19 platform were searched to identify all reports (preprints or peer-reviewed publications) of randomized clinical trials (RCTs) assessing pharmacologic interventions for the treatment of COVID-19, up to May 31, 2021. A standardized, online data-extraction form was developed. Data extraction covered general trial characteristics, transparency indicators (eg, trial registration and data sharing statement), completeness of reporting (eg, 10 of the most important Consolidated Standards of Reporting Trials [CONSORT] 2010 items<sup>3</sup>), and changed outcomes between report and registry (ie, switched outcomes). Two reviewers were trained and assessed 20 trials separately, with an agreement of 96.6% and a  $\kappa$  coefficient of 0.87. A third person resolved disagreements, when needed. A single reviewer assessed all remaining trials. For all trials published first as a preprint, a systematic search was performed for a subsequent publication in peer-reviewed journals, up to October 7, 2021. In a second step, the peer-reviewed journal publication was assessed and compared with the matched preprint.

**Results** A total of 251 trial reports were identified: 121 peer-reviewed journal publications (48%) and 130 preprints (52%). Approximately half of the trials were prospectively registered (140 [56%]); 38% ( $n = 95$ ) made their full protocols or statistical analysis plan available, and 29% ( $n = 72$ ) provided access to their statistical analysis plan report. A data sharing statement was available in 68% of the reports ( $n = 170$ ); 91% stated their willingness to share. Only 32% of the trials ( $n = 81$ ) completely defined the prespecified primary outcome measures; 57% ( $n = 143$ ) reported the process of allocation concealment. Overall, 51% ( $n = 127$ ) adequately reported results for the primary outcomes, whereas only 14% of the trials ( $n = 36$ ) adequately described harms. Primary outcome(s) reported in trial registries and published reports were inconsistent in 49% of the trials ( $n = 104$ ); only 15% ( $n = 16$ ) disclosed outcome switching in the report. There were no major differences between preprints and peer-reviewed publications. Of the 130 RCTs published as a preprint, 78 were then published in a peer-reviewed journal (median delay

to journal publication, 94 days; IQR, 55–168 days). There was no major improvement after the peer review process.

**Conclusions** Lack of transparency, completeness, and consistency of reporting are important barriers to trust, interpretation, and synthesis in COVID-19 clinical trials. Peer-reviewed publications did not report the items assessed better than preprints. The comparison of paired reports published as a preprint and as a peer-reviewed publication did not indicate major improvements.

### References

- Else H. How a torrent of COVID science changed research publishing—in 7 charts. *Nature*. 2020;588(7839):553. doi:10.1038/d41586-020-03564-y
- Besançon L, Peiffer-Smadja N, Segalas C, et al. Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med Res Methodol*. 2021;221(1):117. doi:10.1186/s12874-021-01304-y
- Chauvin A, Ravaud P, Moher D, et al. Accuracy in detecting inadequate research reporting by early career peer reviewers using an online CONSORT-based peer-review tool (COBPeer) versus the usual peer-review process: a cross-sectional diagnostic study. *BMC Med*. 2019;17(1):205. doi:10.1186/s12916-019-1436-0

<sup>1</sup>Université de Paris, INSERM, INRAE, CNAM, Centre de Recherche en Épidémiologie et Statistiques, Paris, France, kapp@ifem.uni-freiburg.de; <sup>2</sup>Centre d’Épidémiologie Clinique, AP-HP, Hôpital Hôtel-Dieu, Paris, France; <sup>3</sup>Cochrane France, Paris, France; <sup>4</sup>Institute for Evidence in Medicine (for Cochrane Germany Foundation), Medical Center University of Freiburg, Freiburg, Germany

**Conflict of Interest Disclosures** Isabelle Boutron is a member of the Peer Review Congress Advisory Board but was not involved in the editorial review or decision for this abstract. No other disclosures were reported.

**Funding/Support** This study was part of the COVID-NMA initiative, which received funding from Université de Paris, Assistance Publique Hôpitaux de Paris, INSERM, Cochrane France (Ministry of Health), the French Ministry of Higher Education and Research, Agence Nationale de la Recherche, and the World Health Organization.

### Day and Time of Submissions of Manuscripts to the *Journal of Paediatrics and Child Health* Before and During the COVID-19 Pandemic

Richard G. McGee,<sup>1,2</sup> Lara E. Graves<sup>3</sup>

**Objective** To determine whether COVID-19 restrictions reduced manuscript submissions during the standard working day.

**Design** This was a retrospective cohort study of all submissions from January 2015 to March 2022 to the *Journal of Paediatrics and Child Health*, a bimonthly peer-reviewed medical journal and the official journal of the Royal Australasian College of Physicians' Paediatrics and Child Health Division. The submission time stamps for all manuscripts submitted by Australia-based authors for which

a final editorial decision had been recorded were analyzed. The first confirmed case of COVID-19 in Australia was identified in January 2020, and the first COVID-19-related restrictions were introduced to some Australian states in March 2020, so the data set was divided into 2 time periods: before COVID-19 (January 2015 to February 2020 inclusive) and the COVID-19 era (March 2020 to March 2022 inclusive). Two-tailed *t* tests were performed to determine the differences in proportions between the time periods for the standard working week vs weekends and between standard working hours (9 AM to 5 PM) vs early hours (12 to 8 AM) vs late hours (6 to 11 PM).

**Results** There were a total of 8244 manuscripts submitted to the journal during this period. Of these, 2167 manuscripts were submitted by Australia-based researchers, and there were 2114 manuscripts for which an editorial decision was recorded (53 under review). A median of 286 manuscripts were submitted by Australia-based researchers each year. **Table 83** shows the day and time of manuscript submission. Before the onset of COVID-19, 89.4% of submissions were made during the standard working week (Monday through Friday); during the COVID-19 era, this proportion reduced to 80.5%, although the difference was not statistically significant ( $P = .48$ ). Before the onset of COVID-19, 67.3% of submissions were made during standard working hours (9 AM to 5 PM); during the COVID-19 era, this reduced to 61.5%, but the difference was not statistically significant ( $P = .43$ ). Likewise, the differences in proportion of submissions for the time periods 12 to 8 AM (14.2% vs 16.7%;  $P = 0.7$ ) and 6 to 11 PM (18.5% vs 21.8%;  $P = .40$ ) were not statistically significant.

**Conclusions** Most manuscript submissions occurred during the standard working week and working hours, with no significant change before and after COVID-19 restrictions. The day and time of manuscript submission represents only one aspect of the processes involved in conducting research and may not correlate to when other parts of research are performed. However, COVID-19 restrictions are ongoing in Australia, and their effect on research practices may still be evolving.

<sup>1</sup>Department of Paediatrics, Gosford Hospital, Gosford, New South Wales, Australia; <sup>2</sup>Central Coast Clinical School, The University of Newcastle, Gosford Hospital, Gosford, New South Wales, Australia, richard.mcgee@newcastle.edu.au; <sup>3</sup>Department of Paediatric Endocrinology, The Children's Hospital at Westmead, Westmead, New South Wales, Australia

**Acknowledgments** We thank Ms Kate Edmonds and Prof David Isaacs for their assistance in obtaining the editorial data.

### Assessing Repeated Patient Information in Systematic Reviews Published Early in the COVID-19 Pandemic

Pablo J. Moreno-Peña,<sup>1</sup> Miguel Zambrano-Lucio,<sup>1</sup> Francisco J. Barrera,<sup>1,2,3</sup> Andrea Flores Rodríguez,<sup>1</sup> Skand Shekhar,<sup>4</sup> Rachel Wurth,<sup>5</sup> Michelle Hajdenberg,<sup>6</sup> Neri A. Alvarez-Villalobos,<sup>1,2,3,7</sup> Janet E. Hall,<sup>4</sup> Ernesto L. Schiffrin,<sup>8</sup> Juan P.

**Table 83. Day and Time of Article Submission**

Day and time	Submissions, No. (%)		<i>P</i> value <sup>a</sup>
	January 2015 to February 2020 (n = 1426)	March 2020 to March 2022 (n = 688)	
<b>Day</b>			
Monday	244 (17.1)	109 (18.7)	.48
Tuesday	267 (18.7)	124 (18.1)	
Wednesday	262 (18.4)	124 (17.7)	
Thursday	264 (18.5)	145 (18.9)	
Friday	238 (16.7)	90 (7.1)	
Saturday	73 (5.1)	51 (7.9)	
Sunday	78 (5.5)	45 (11.5)	.24
<b>Time</b>			
12:00 AM	6 (0.4)	7 (1.0)	.70
1:00 AM	2 (0.1)	7 (1.0)	
2:00 AM	5 (0.4)	8 (1.2)	
3:00 AM	4 (0.3)	7 (1.0)	
4:00 AM	8 (0.6)	4 (0.6)	
5:00 AM	29 (2.0)	12 (1.7)	
6:00 AM	40 (2.8)	16 (2.3)	
7:00 AM	43 (3.0)	23 (3.3)	
8:00 AM	66 (4.6)	31 (4.5)	
9:00 AM	125 (8.8)	38 (5.5)	.43
10:00 AM	108 (7.6)	60 (8.7)	
11:00 AM	136 (9.5)	44 (6.4)	
12:00 PM	122 (8.6)	42 (6.1)	
1:00 PM	117 (8.2)	53 (7.7)	
2:00 PM	105 (7.4)	69 (10.0)	
3:00 PM	97 (6.8)	48 (7.0)	
4:00 PM	90 (6.3)	36 (5.2)	
5:00 PM	59 (4.1)	33 (4.8)	
6:00 PM	49 (3.4)	36 (5.2)	.40
7:00 PM	56 (3.9)	30 (4.4)	
8:00 PM	60 (4.2)	30 (4.4)	
9:00 PM	39 (2.7)	22 (3.2)	
10:00 PM	32 (2.2)	17 (2.5)	
11:00 PM	28 (2.0)	15 (2.2)	

<sup>a</sup>*P* values were calculated with 2-tailed *t* tests.

Brito,<sup>2</sup> Stefan R. Bornstein,<sup>9,10,11</sup> Constantine A. Stratakis,<sup>5</sup> Fady Hannah-Shmouni,<sup>5</sup> René Rodríguez-Gutiérrez<sup>1,2,3,7</sup>

**Objective** The inclusion of duplicate publications in systematic reviews (SRs) has led to repeated patient information (RPI).<sup>1,2</sup> Repeated patient information in SRs is the inclusion of a patient's information multiple times, with the assumption that they are different participants. This could result in the overestimation or underestimation of results, which can lead to substantial clinical implications through misleading estimates.<sup>1</sup> A proportion of studies with shared timing and location was identified in SRs by evaluating an

early-stage sample of COVID-19 SRs. According to the International Committee of Medical Journal Editors, a duplicate publication overlaps substantially with a published article without clear reference to the initial publication.<sup>3</sup> The proportion of studies with shared characteristics that could suggest RPI was assessed in this study.

**Design** This study was an umbrella review of SRs with clinical data for patients with COVID-19. An experienced librarian performed a comprehensive search strategy for peer-reviewed articles in the English language published between December 1, 2019, and April 6, 2020, in databases that included Ovid MEDLINE In-Process & Other Non-Indexed Citations, Ovid MEDLINE, Ovid Embase, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus. Studies included in SRs were grouped into clusters depending on characteristics including time frame and location. The frequency of studies that met the definition of at risk of including RPI was measured according to shared timing and location in the same SR.

**Results** Fifteen SRs were included, with a total population of 172,558 participants. A median (IQR) of 9 (6-20) studies were included in each SR, and 6 (2-10) studies included by each SR were considered at risk. Of these, 14 (93.3%) had risk of RPI. Subsequently, 103 clusters were generated. A median (IQR) of 3 (1-6) at-risk clusters were included in each SR. Eleven SRs (73.3%) included articles with RPI from a single hospital.

**Conclusions** Risk of RPI was prevalent in COVID-19 SRs published early in the pandemic, and RPI may also be common in SRs of topics outside COVID-19. The impact of RPI on SRs could dilute their validity. Statements about effect sizes should be made carefully, ensuring studies have carefully selected their population to include unique participants. The following are suggestions to improve the often-complex process of identifying RPI: (1) journals should ask authors to state if any data have been published and, if so, to provide a reference; (2) reporting guidelines (CONSORT, STARD, CARE, STROBE, and PRISMA) should include a domain asking the authors if any of the data has been published elsewhere; and (3) the quality assessment tool for SRs (AMSTAR) should include a domain that evaluates whether the authors evaluate the inclusion of RPI.

## References

1. Choi WS, Song SW, Ock SM, et al. Duplicate publication of articles used in meta-analysis in Korea. *SpringerPlus*. 2014;3(1):182.
2. von Elm E, Poglia G, Walder B, Tramèr MR. Different patterns of duplicate publication: an analysis of articles used in systematic reviews. *JAMA*. 2004;291(8):974-980.
3. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals. International Committee of Medical Journal Editors. Updated May 2022. Accessed January 29, 2022. <http://icmje.org/recommendations/>
- <sup>1</sup>Plataforma INVEST-KER Unit Mayo Clinic (KER Unit Mexico), School of Medicine, Universidad Autonoma de Nuevo Leon, Monterrey, Nuevo León, Mexico, renerodriguezgtz@hotmail.com;
- <sup>2</sup>Knowledge and Evaluation Research, Mayo Clinic, Rochester, MN, USA;
- <sup>3</sup>Endocrinology Division, Internal Medicine Department, Hospital Universitario Dr José Eleuterio González, Universidad Autonoma de Nuevo Leon, Monterrey, Nuevo León, Mexico;
- <sup>4</sup>Clinical Research Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA;
- <sup>5</sup>Section on Endocrinology & Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA;
- <sup>6</sup>College of Arts and Sciences at Washington University in St Louis, St Louis, MO, USA;
- <sup>7</sup>Research Unit, School of Medicine and Hospital Universitario Dr José Eleuterio González, Universidad Autonoma de Nuevo Leon, Monterrey, Nuevo León, Mexico;
- <sup>8</sup>Department of Medicine, Sir Mortimer B. Davis-Jewish General Hospital, McGill University, Montreal, Montreal, QC, Canada;
- <sup>9</sup>Department of Medicine III, University Hospital Carl Gustav Carus, Dresden, Germany;
- <sup>10</sup>Department of Diabetes, School of Life Course Science & Medicine, King's College London Strand, London, UK;
- <sup>11</sup>Klinik für Endokrinologie, Diabetologie und Klinische Ernährung, University Hospital of Zürich, Zürich, Switzerland

## Conflict of Interest Disclosures

**Funding/Support** This study was funded in part by the intramural research program of the National Institutes of Health.

**Role of the Funder/Sponsor** The sponsor was not involved in the design or conduct of the study, the preparation of the abstract, or the decision to submit the abstract.

**Additional Information** Pablo J. Moreno-Peña and Miguel Zambrano-Lucio share equal credit as co-first authors. Fady Hannah-Shmouni is a co-corresponding author.

## Peer Review

### An International Survey of Biomedical Researchers' Knowledge, Perceptions, and Training on Peer Review

Jess V. Willis,<sup>1,2</sup> Kelly D. Cobey,<sup>3,4</sup> Janina Ramos,<sup>1,5</sup> Mohsen Alayche,<sup>1,2</sup> Jeremy Y. Ng,<sup>1</sup> David Moher<sup>1,4</sup>

**Objective** To provide an up-to-date perspective of biomedical researchers' knowledge and perceptions of and engagement with peer review training. Previous studies, many done in consultation with editors and publishers, have reported researchers' attitudes on peer review.<sup>1,2</sup> This survey study is descriptive and does not have any *a priori* hypotheses.

**Design** A cross-sectional online survey of biomedical researchers around the world was conducted. The CHERRIES (Checklist for Reporting Results of Internet E-Surveys) reporting guideline was used to inform the reporting of findings. A random sample of 2000 corresponding authors was collected from a Scopus source list of recently published articles in peer-reviewed biomedical journals. Authors from journals that exclusively published non-English articles were excluded. The closed survey was purpose built for this study, was administered using SurveyMonkey, and was available only to participants identified via the random sampling

framework. Participants were invited through email; participation was voluntary, and all data were collected anonymously. Collected data included participants' demographic characteristics as well as their experience with and opinions about peer review, with additional open-ended questions allowing participants to elaborate their responses. Data were analyzed from all surveys in which participants responded to 80% or more of the questions.

**Results** Of the 2000 invited researchers, 186 (9.3%) responded. The average survey completion rate among these participants was 91%. Most participants (142 [76.3%]) reported having 6 or more years of experience in scholarly publishing. One hundred two of 180 participants (56.7%) reported being active as a manuscript peer reviewer for more than 6 years, and 171 of 185 participants (92.4%) reported having peer reviewed at least 1 article in the last 12 months. Despite the robust experience and activity in manuscript peer review reported by the participants, only 28 of 185 participants (15.1%) completed formal training in peer review. Twelve of 64 participants (18.8%) received training through in-person lecture, and 11 of 64 participants (17.2%) received training through online lecture. Thirteen of 36 participants (52.8%) received training in peer review provided by a university or college.

**Conclusions** This study will provide a current international perspective on biomedical researchers' knowledge, perceptions, and engagement regarding peer review training. The results of the survey may help identify gaps in peer review training experience and knowledge. Subsequently, the findings may guide the creation of future training options, inform the development of preferred training methods, and increase comprehensiveness of peer review training for biomedical researchers.

## References

1. Mulligan A, Hall L, Raphael E. Peer review in a changing world: an international study measuring the attitudes of researchers. *J Am Soc Inf Sci Tec.* 2013;64:132-161. doi:10.1002/asi.22798
2. Publons. 2018 Global state of peer review. 2018. Accessed July 2022. <https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf>

<sup>1</sup>Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada, jramo101@uottawa.ca; <sup>2</sup>Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>3</sup>University of Ottawa Heart Institute, Ottawa, ON, Canada; <sup>4</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; <sup>5</sup>Department of Biology, Faculty of Science, University of Ottawa, Ottawa, ON, Canada

**Conflict of Interest Disclosures** David Moher is an associate director of the International Congress on Peer Review and Scientific Publication but was not involved in the review or decision for this abstract. No other disclosures reported.

## Development of a List to Detect Statistical and Methodological Terms in Peer Reviews

Ivan Buljan,<sup>1</sup> Daniel Garcia-Costa,<sup>2</sup> Francisco Grimaldo,<sup>2</sup> Richard A. Klein,<sup>3</sup> Marjan Bakker,<sup>3</sup> Ana Marušić<sup>1</sup>

**Objective** The peer review system used by academic journals to assess the quality of submitted papers is constantly in need of new technologies that can help reviewers and increase the objectivity of the process. The aim of the present study was to develop a comprehensive list of words related to research methods and statistics, which could then be used to automatically study the content and the quantity and types of methodological and statistical issues identified in peer review reports.

**Design** The terms for the list were identified from different glossaries.<sup>1,2</sup> The terms were organized into subcategories, some of which were more related to methodological aspects (eg, study design, sampling, or procedures) or statistical aspects (eg, parametric descriptive parameters or data presentation) of research. The list was refined and enriched with terms using a computational language model. The final list and the preregistration can be found at the Open Science Framework website (<https://osf.io/d34b9/>). The occurrence of terms in the PEERE database<sup>3</sup> of peer review reports, for which data on reviewer gender, continental region, review recommendation, journal impact factor, and subject area (N = 496,928) were available, was counted. Regression models were created to determine the frequency of the terms used based on peer review characteristics.

**Results** The mean length of a peer review was 196 words (95% CI, 108-330 words). A total of 26.8% of review reports in the sample contained at least 1 methodological term, while 70.7% contained at least 1 statistical term. It was more likely that the review contained methodological terms if the reviewer was female and resided in North America or Oceania compared with other continents, if the recommendation was to reject the article or send it for major revisions, if the journal was from the social sciences and humanities, if the journal was in the top impact factor quartile, and if the article was a longer peer review. More words related to statistics were observed when the reviewer was female, for reviews recommending any revisions, for journals from health and medical sciences, in journals in the top impact factor quartile, and in longer reviews. Mixed regression models indicated that the scientific area was more strongly associated with words related to statistics than were reviewer recommendations.

**Conclusions** The application of this newly developed list showed that methodological and statistical terms are not a common topic in peer review and that they are not always present in review reports from different fields. The list successfully covered research terms because the most-identified terms were related to categories relevant for most types of research. Future analysis should try to understand the context of the occurrence of words from the list to confirm the validity of the list of terms.

## References

1. Stark PB. Glossary of Statistical Terms. Accessed March 23, 2021. <https://www.stat.berkeley.edu/~stark/SticiGui/Text/gloss.htm>
2. Framework for Open and Reproducible Research Training. The Glossary of Terms. Accessed July 12, 2021. <https://docs.google.com/document/d/1oNsHSRmgAsDGUKvq5joqBpvHeRmDKldH2kSyJovHhk/edit#heading=h.2l8ssl89e1ey>
3. Squazzoni F, Ahrweiler P, Barros T, et al. Unlock ways to share data on peer review. *Nature*. 2020;578(7796):512-514. doi:10.1038/d41586-020-00500-y

<sup>1</sup>Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia, ibuljan@mefst.hr;

<sup>2</sup>Department of Computer Science, University of Valencia, Burjassot, Spain; <sup>3</sup>Department of Methodology and Statistics, Tilburg University, Tilburg, the Netherlands

**Conflict of Interest Disclosures** Ana Marušić is a member of the Peer Review Congress Advisory Board by was not involved in the editorial review or decision for this abstract. No other disclosures reported.

**Funding/Support** This study was funded by the Croatian Science Foundation “Professionalism in Health–Decision making in practice and research” under grant agreement IP-2019-04-4882. Richard A. Klein received funding from a Replication Studies grant (401.18.053) from the Netherlands Organization for Scientific Research, and a Consolidator Grant (IMPROVE) from the European Research Council (grant 726361). Francisco Grimaldo and Daniel García-Costa are partially supported by the Spanish Ministry of Science, Innovation and Universities, the Spanish State Research Agency, and the European Regional Development Fund under project RTI2018-095820-B-I00.

**Role of the Funder/Sponsor** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

## Publication Metrics and Performance Indicators

### Sorting Out Journals: Quality Criteria, Ranking Principles, and Tensions of Chinese Scientific Journal Lists

Jing Wang,<sup>1</sup> Willem Halfman,<sup>1</sup> Yuehong Zhang<sup>2</sup>

**Objective** Journal lists are developed in China to avoid the domination of the journal impact factor in journal and research evaluation. Journal lists are instruments to categorize, compare, and assess scholarly publications. Avoiding the misleading precision of indicators, these simpler ordinal or nominal lists have been established, evaluated, used, and debated by different users of scholarly publishing channels globally. This study investigated the remarkable proliferation of journal lists in China and analyzed their underlying values, quality criteria, and ranking principles. In contrast with well-established international lists, this study investigated the concerns specific to the Chinese research policy and publishing system.

**Design** This qualitative study was based on an analysis of policy documents concerning Chinese research and publishing policy and specific list-making initiatives. This study investigated 20 disqualifying journal lists (2 authoritative lists and 18 lists from universities and hospitals) and 2 qualifying lists (**Table 84**). The document analysis was complemented by interviews with journal list makers to investigate the list-making process. The study focused on Chinese journals in science, technology, engineering, and mathematics (STEM).

**Results** In an overview of the current Chinese journal lists, several key distinctions and contrasts were highlighted in listing criteria. Disqualifying lists of “bad journals” reflect concerns over inferior research publications, but also the involved drain on public resources. For example, the National Science Library of the Chinese Academy of Sciences uses 7

**Table 84. Scientific Journal Lists in China**

Journal lists		Sponsoring organization	No. of journals	Quality differentiation levels	Purpose
List of discouraged journals	Early warning journal list	National Science Library of Chinese Academy of Sciences	65 in 2020 and 36 in 2021	3 Risk levels	Warning researchers and publishers
	Warning and discourage list	Institute of Scientific and Technical Information of China	91	2 Warning levels	Guiding administration/funding departments of universities or research institutes
	Negative journal list	8 Universities and 10 hospitals	7-60	Some lists with differentiated levels and some without	Publishing expense claims and awards
Lists of endorsed journals	Chinese STEM Journal Excellence Action Plan	7 Ministries	280 in 2019	4 Categories	Funding allocation to good journals for development
	High-quality STEM journal catalogue graded by field	Funded by the Chinese Association for Science and Technology; implemented by different academic societies	Different based on academic societies	3 Tiers	To award Chinese journals the same credits as international journals in terms of faculty promotion and merit system

Abbreviation: STEM, science, technology, engineering, and mathematics.

criteria to compile the *Early Warning List of International Journals* to inform researchers' publishing choices and publishers' journal quality management. Qualifying lists of "good journals" are based on criteria valued in research policy, typically sorting journals into ordinal quality levels. The considerations in the development of these lists reflect specific policy concerns. For example, the *Chinese STM Journal Excellence Action Plan* generated a journal funding list as a reference for public investment in journals. The *High-Quality STEM Journal Catalogue Graded by Fields* includes evaluative lists of domestic and international journals for use in academic evaluation. Contrasting concerns and inaccuracies lead to contradictions in the qualify and disqualify binary logic, as demonstrated in the case of a journal listed on both the qualifying and disqualifying lists. Similarly, different qualifying lists provide different assessments of what constitutes a good or excellent journal.

**Conclusions** The administrative logic of state-led Chinese research and publishing policy ascribes worth to scientific journals for its specific national and institutional needs. These needs involve the challenges of public resource allocation, a shift away from output dominated research evaluation, research misconduct, and balancing national research needs against international standards. Therefore, Chinese journal lists use quality criteria in a specific way that is different from other journal lists. However, journal lists may not always be able to represent both general journal quality and quality for specific purposes.

## References

1. Pölonen J, Guns R, Kulczycki E, Sivertsen G, Engels TCE. National lists of scholarly publication channels: an overview and recommendations for their construction and maintenance. *J Data Inf Sci.* 2021;6(1):50. doi:10.2478/jdis-2021-0004
2. Zhang L, Sivertsen G. The new research assessment reform in China and its implementation. *Scholarly Assessment Reports.* 2020;2(1):7. doi:10.29024/sar.15
3. Fochler M, Felt U, Müller R. Unsustainable growth, hyper-competition, and worth in life science research: narrowing evaluative repertoires in doctoral and postdoctoral scientists' work and lives. *Minerva.* 2016;54(2):175–200. doi:10.1007/s11024-016-9292-y

<sup>1</sup>Institute for Science in Society, Radboud University, Nijmegen, the Netherlands, jing.wang@ru.nl; <sup>2</sup>Journal of Zhejiang University-Science, Zhejiang University, Hangzhou, China

**Conflict of Interest Disclosures** None reported.

**Funding/Support** The work reported in this abstract is funded by the China Scholarship Council.

## Quality of the Literature

### Characteristics and Opportunities for Improvement of Methods Guidance Published in General and Methodology-Focused Medical Journals

Julian Hirt,<sup>1,2</sup> Hannah Ewald,<sup>3</sup> Daeria O. Lawson,<sup>4</sup> Lars G. Hemkens,<sup>1,5,6</sup> Matthias Briel,<sup>1,4</sup> Stefan Schandelmaier<sup>1,4</sup>

**Objective** To systematically assess the characteristics and current practice of developing methods guidance for health researchers and explore opportunities for improvement.

**Design** A systematic survey of methods-guidance articles published in general and methodology-focused, high-impact medical journals indexed in MEDLINE in 2020 was performed. Articles that explicitly stated the objective to provide methods guidance for health research were eligible. Characteristics related to findability, methods for guidance development, and transparency were extracted.

**Results** A total of 105 guidance articles published in 12 different journals were included. Few articles had a structured abstract (44 [42%]) or were indexed with Medical Subject Headings (40 [38%]) or author keywords (18 [17%]) related to guidance. Of the 105 guidance articles, less than half reported any methods for development (44 [42%]), most frequently stakeholder involvement (30 [28%]), systematic review of the methodological literature (21 [20%]), or consensus process (21 [20%]). Use of explicit methods for development differed between reporting guidelines (13 of 13 [100%] reported a development process) and other types of methods guidance (ie, guidance for planning, conduct, analysis, interpretation, or quality assessment: 31 of 92 articles [34%] reported a development process). Transparency was limited, with few guidance articles describing the authors' expertise (23 [22%]). Conflicts of interest, if reported (36 [34%]), were frequently unclear.

**Conclusions** Most methods-guidance articles published in 2020 were difficult to find in MEDLINE, were developed with unclear methods, and lacked transparency regarding the authors' expertise and conflicts of interest. For health researchers, those limitations implied important barriers to the uptake of methods guidance. To improve findability, we developed the new open-access Library of Guidance for Health Scientists (LIGHTS; <https://lights.science>). More research is required to inform methods for guidance development and transparency considerations.

<sup>1</sup>Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland, s.schandelmaier@gmail.com; <sup>2</sup>International Graduate Academy, Institute for Health and Nursing Science, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany; <sup>3</sup>University Medical Library, University of Basel, Basel, Switzerland; <sup>4</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada; <sup>5</sup>Meta-Research Innovation Center Berlin (METRIC-B), Berlin Institute of Health, Berlin, Germany; <sup>6</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

**Conflict of Interest Disclosures** None reported.

**Funding/Support** This work was supported by grant 190566 from the Swiss National Science Foundation and grant 3MS1043 from the University of Basel.

**Role of the Funder/Sponsor** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.

**Additional Information** Julian Hirt is a co-corresponding author.

### Comparison of Changes in High-Quality vs Low-Quality Evidence in Original and Updated Systematic Reviews

Benjamin Djulbegovic,<sup>1</sup> Muhammad Muneeb Ahmed,<sup>2</sup> Iztok Hozo,<sup>3</sup> Despina Koletsis,<sup>4</sup> Lars Hemkens,<sup>5,6,7</sup> Amy Price,<sup>8</sup> Rachel Riera,<sup>9</sup> Paulo Nadanovsky,<sup>10</sup> Ana Paula Pires dos Santos,<sup>11</sup> Daniela Melo,<sup>12</sup> Ranjan Pathak,<sup>13</sup> Rafael Leite Pacheco,<sup>14,15</sup> Luis Eduardo Fontes,<sup>15,16</sup> Enderson Miranda,<sup>17</sup> David Nunan,<sup>17,18</sup>

**Objective** It is generally believed that evidence from a weaker body of evidence (eg, poorly designed and executed and potentially biased studies, sparse studies and/or heterogenous results) will generate inaccurate estimates about treatment effects more often than evidence from a stronger body of evidence. As a result, estimates of effects of health interventions initially based on high certainty (quality) of evidence (CoE) are expected to change less frequently than the effects estimated by lower CoE, and the estimates of magnitude of effect size are expected to differ between high and low CoE. Empirical assessment of these foundational principles of evidence-based medicine has been lacking.

**Design** The Cochrane Database of Systematic Reviews was reviewed from January 2016 through May 2021 for pairs of original and updated reviews for change in CoE assessments based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) method. The difference in effect sizes between the original and updated reviews were assessed as a function of change in CoE, which was reported as a ratio of odds ratios (RORs). The RORs generated in the studies that changed CoE from very low/low (VL/L) to moderate/high (M/H) vs MH/H to VL/L were compared. Heterogeneity and inconsistency were assessed using the  $\tau^2$  and  $I^2$  statistic. The change in precision of effect size estimates was assessed by calculating the ratio of standard errors (seR), and the absolute deviation in estimates of treatment effects was assessed with adjusted RORs.

**Results** Overall, 419 pairs of reviews were included, of which 414 (207  $\times$  2) informed the CoE appraisal and 384 (192  $\times$  2) the assessment of effect size. Certainty of evidence originally appraised as VL/L had 2.1 (95% CI, 1.19-4.12;  $P = .01$ ) times higher odds to be changed in the future studies than those with M/H CoE. However, the pooled effect size was not different when the CoE changed from VL/L to M/H (ROR, 1.02; 95% CI, 0.74-1.39) compared with M/H CoE changing

to VL/L (ROR, 1.02; 95% CI, 0.44-2.37). Similarly, the overlap in aROR between the VL/L CoE to M/H vs the M/H to VL/L subgroups was observed (median [IQR], 1.12; 95% CI, 1.07-1.57 vs 1.21; 95% CI, 1.12-2.43). There was a large inconsistency across ROR estimates ( $I^2 = 99\%$ ). There was larger imprecision in treatment effects when the CoE changed from VL/L to M/H (seR = 1.46) than when it changed from M/H to VL/L (seR = 0.72).

**Conclusions** This study found that low-quality evidence changed more often than high CoE. However, the effect size was not systematically different between studies with low vs high CoE, indicating the need for improving contemporary critical appraisal methods.

<sup>1</sup>Beckman Research Institute, Department of Computational & Quantitative Medicine, City of Hope, Duarte, CA, USA, bdjulbegovic@coh.org; <sup>2</sup>Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada; <sup>3</sup>Department of Mathematics, Indiana University Northwest, Gary, IN, USA; <sup>4</sup>Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland; <sup>5</sup>Department of Clinical Research, University of Basel, Basel Institute for Clinical Epidemiology & Biostatistics, University Hospital Basel, Basel, Switzerland; <sup>6</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA; <sup>7</sup>Meta-Research Innovation Center Berlin (METRIC-B), Berlin Institute of Health, Berlin, Germany; <sup>8</sup>Anesthesia Informatics and Media Lab, Stanford University, Stanford, CA, USA; <sup>9</sup>Universidade Federal de São Paulo, Escola Paulista de Medicina, Brazil (Unifesp), São Paulo, Brazil; <sup>10</sup>Department of Epidemiology and Quantitative Methods in Health, National School of Public Health, Fundação Oswaldo Cruz (FIOCRUZ)–Department of Epidemiology, Institute of Social Medicine, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, Brazil; <sup>11</sup>Department of Community and Preventive Dentistry, Faculty of Dentistry, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, Brazil; <sup>12</sup>Department of Pharmaceutical Sciences, Universidade Federal de São Paulo (Unifesp), São Paulo, Brazil; <sup>13</sup>Department of Medical Oncology and Therapeutics Research, City of Hope, Duarte, CA, USA; <sup>14</sup>Centro Universitário São Camilo, São Paulo, Brazil; <sup>15</sup>Center of Health Technology Assessment, Hospital Sírio-Libanês–São Paulo, Brazil; <sup>16</sup>Department of Intensive Care and Emergency Medicine, Faculdade de Medicina de Petrópolis, Petrópolis, Rio de Janeiro, Brazil; <sup>17</sup>Kellogg College, University of Oxford, Oxford, UK; <sup>18</sup>Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences, Oxford University, Oxford, UK

**Conflict of Interest Disclosures** None reported.

## Quality of Reporting

### Completeness of Reporting and Its Association With Risk of Bias in Systematic Reviews Published in Rehabilitation Journals: A Meta-research Study

Tiziano Innocenti,<sup>1,2</sup> Daniel Feller,<sup>3</sup> Silvia Giagio,<sup>4,5</sup> Stefano Salvioli,<sup>2,6</sup> Silvia Minnucci,<sup>7</sup> Fabrizio Brindisino,<sup>7,8</sup> Carola Cosentino,<sup>6</sup> Leonardo Piano,<sup>9</sup> Alessandro Chiarotto,<sup>1,10</sup> Raymond Ostelo<sup>1,11</sup>

**Objective** Problems with reporting a scientific study can affect research in different manners. For example, it is known that study methods are frequently not described in adequate

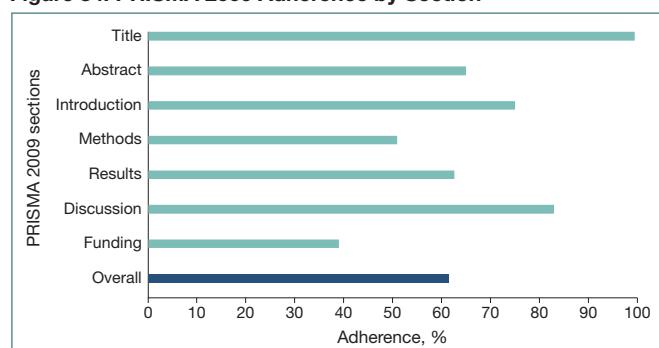
detail and that results are presented ambiguously, incompletely, or selectively. A previous study<sup>1</sup> confirmed that most of the authors (approximately 67%) publishing systematic reviews (SRs) in high-impact rehabilitation journals did not mention the use of a reporting guideline. Therefore, the objectives of this study were (1) to evaluate the completeness of reporting of SRs published in rehabilitation journals by evaluating their adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 checklist, (2) to investigate the relationship between completeness of reporting and risk of bias (ROB), and (3) to study the association between completeness of reporting and the characteristics of SRs and journals.

**Design** A random sample of 200 SRs published between 2011 and 2020 in the 68 rehabilitation journals was indexed under the rehabilitation category in the InCites Journal Citation Report. Two independent reviewers evaluated the completeness of reporting as the adherence to the PRISMA 2009 checklist and assessed ROB using the ROBIS tool. Overall adherence and adherence to each PRISMA 2009 item and section were calculated as a percentage of the total number of items described and reported of the total number of applicable items. Regression analyses investigating the association between completeness of reporting, ROB, and other characteristics (impact factor, publication options, year of publication, and study protocol registration) were performed.

**Results** The mean overall PRISMA 2009 adherence was 61.4% (**Figure 34**). Studies with high ROB in ROBIS domain 1 had an overall adherence to the PRISMA 2009 checklist, which was 5.70% (95% CI, -10.07% to -1.34%) lower than those with low ROB. In domain 2, studies with high ROB had 5.41% (95% CI, -9.74% to -1.07%) lower adherence than those with low ROB. Studies with high overall ROB had 7.06% (95% CI, -12.10% to -2.01%) lower adherence than those with low overall ROB. Studies published in fourth-quartile journals displayed an overall adherence of 7.24% (95% CI, -13.19% to -1.29%) lower than those published in the first quartile; there was an 11.95% (95% CI, 5.94%-17.96%) increase in overall adherence if the SR protocol was registered.

**Conclusions** The completeness of reporting in SRs published in rehabilitation journals is still suboptimal. High ROB is associated with poorer completeness of reporting.

**Figure 34. PRISMA 2009 Adherence by Section**



indicating that if a study has better reporting, this could positively affect the transparency of the ROB evaluation. The registration of study protocols and journal ranking are also associated with a more complete reporting. Authors of SRs should improve adherence to the reporting standards, and journal editors could implement strategies to optimize the completeness of reporting.

## Reference

- Innocenti T, Salvioli S, Giagio S, Feller D, Cartabellotta N, Chiarotto A. Declaration of use and appropriate use of reporting guidelines in high-impact rehabilitation journals is limited: a meta-research study. *J Clin Epidemiol*. 2021;131:43-50. doi:10.1016/j.jclinepi.2020.11.010

<sup>1</sup>Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, the Netherlands; <sup>2</sup>GIMBE Foundation, Bologna, Italy;

<sup>3</sup>Provincial Agency for Health of the Autonomous Province of Trento, Trento, Italy; <sup>4</sup>Division of Occupational Medicine, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy;

<sup>5</sup>Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum, University of Bologna, Bologna, Italy;

<sup>6</sup>Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genoa, Italy; <sup>7</sup>Department of Clinical Science and Translational Medicine, University of Rome Tor Vergata, Rome, Italy; <sup>8</sup>Department of Medicine and Health Science "Vincenzo Tiberio," University of Molise, Campobasso, Italy; <sup>9</sup>Unit of Rehabilitation and Functional Recovery, Fondazione dei Santi Lorenzo e Teobaldo, Rodello, Italy; <sup>10</sup>Department of General Practice, Erasmus MC, University Medical Center, Rotterdam, the Netherlands; <sup>11</sup>Department of Epidemiology and Data Science, Amsterdam UMC, Location VUmc, Amsterdam Movement Sciences Research Institute, Amsterdam, the Netherlands

**Conflict of Interest Disclosures** None reported.

## Methodological and Reporting Quality of Systematic Reviews in Dermatology

Annapoorani Muthiah,<sup>1</sup> Loch Kith Lee,<sup>1</sup> John Koh,<sup>1</sup> Ashley Liu,<sup>1</sup> Aidan C. Tan<sup>2</sup>

**Objective** Over the past decade, there has been an exponential increase in systematic reviews and meta-analyses (SR/MAs) published in the medical literature, and some concerns have arisen that many of these are duplicate studies with redundant results or new studies of unclear quality.<sup>1-3</sup> The aim of this study was to determine whether there was any change in the methodological or reporting quality of SRs in dermatology over the past decade.

**Design** This was a cross-sectional study of all SR/MAs published in 2010 and 2019 in the 10 highest-ranked dermatology journals by SCImago Journal. Methodological quality was assessed through duplicate and independent adjudication by 2 reviewers (A.M., L.K.L., J.K., A.L.) using the Risk of Bias in Systematic reviews (ROBIS) tool and, for SR/MA of interventions, the A Measurement Tool to Assess Systematic Reviews (AMSTAR) 2 tool. Reporting quality was assessed by 1 investigator (A.M.) using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and PRISMA for Abstracts (PRISMA-A) tools, and

study, author, and journal characteristics were abstracted. Statistical analysis was performed with Jamovi, version 1.6.23. Methodological quality was analyzed with the  $\chi^2$  test for independent proportions, or the Fisher exact test if any of the expected frequencies were less than 5, and difference in proportions with 95% CIs. Reporting quality was analyzed with the independent samples  $t$  test and difference in means with 95% CIs.

**Results** In total, 21 SR/MAs from 2010 and 127 from 2019 were included. There was little to no difference between 2010 and 2019 in the proportion of SR/MAs at high or unclear overall risk of bias with ROBIS or with critically low methodological quality using AMSTAR 2. The only subdomain of ROBIS with a difference in proportion of SR/MAs at high or unclear risk of bias between 2010 and 2019 was eligibility bias, with 27.3% more (95% CI, 5.42%-49.2%) in 2010 (66.7%) than 2019 (39.4%) ( $P = .02$ ). There was a strong difference in proportion of PRISMA ( $t_{146} = 3.15$ ;  $P = .002$ ) and PRISMA-A ( $t_{146} = 2.46$ ;  $P = .02$ ) checklist items adequately reported between 2010 and 2019. The difference in mean proportion of PRISMA checklist items adequately reported was 3.8 items more (95% CI, 1.4-6.2 items more) in 2019 (mean [SD], 15.7 [5.1] items) than in 2010 (mean [SD], 11.9 [5.5]), and of PRISMA-A checklist items adequately reported was 0.9 items more (95% CI, 0.2-1.6 items more) in 2019 (mean [SD], 5.9 [1.5] items) than in 2010 (mean [SD], 5 [1.9] items).

**Conclusions** While there was an improvement in the overall reporting quality of SR/MAs between 2010 and 2019, there was no improvement in the overall methodological quality as assessed by the ROBIS and AMSTAR 2 tools. This suggests a persistent potential for forming unreliable conclusions in SR/MAs through a decade of dermatological research. These results therefore highlight the urgency with which efforts to improve SR/MA methodology should be undertaken.

## References

1. Siontis KC, Ioannidis JPA. Replication, duplication, and waste in a quarter million systematic reviews and meta-analyses. *Circ Cardiovasc Qual Outcomes*. 2018;11(12):e005212.
2. Katsura M, Kuriyama A, Tada M, et al. High variability in results and methodological quality among overlapping systematic reviews on the same topics in surgery: a meta-epidemiological study. *Br J Surg*. 2021;108(12):1521-1529.
3. Gómez-García F, Ruano J, Gay-Mimbrera J, et al. Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool. *J Clin Epidemiol*. 2017;92:79-88.

<sup>1</sup>University of New South Wales Medicine, University of New South Wales, Sydney, New South Wales, Australia; <sup>2</sup>National Health and Medical Research Council Clinical Trials Center, University of Sydney, Sydney, New South Wales, Australia, aidan.tan@sydney.edu.au

**Conflict of Interest Disclosures** None reported.

## Registries and Repositories

### Factors Affecting Publication of Pediatric Intervention Trials

Sumaira Khalil,<sup>1</sup> Devendra Mishra,<sup>2</sup> Dheeraj Shah<sup>1</sup>

**Objective** To assess the publication status and factors associated with subsequent publication of all pediatric intervention trials registered in the Clinical Trial Registry of India (CTRI) over a period of 5 years.

**Design** This cross-sectional study was conducted from December 2021 to February 2022 and included the first 100 pediatric intervention trials registered in the CTRI from 2008 to 2012. Registry records were identified from the CTRI website using the keywords *pediatric, paediatric, children, adolescent, infant, newborn, neonate, kids, and school*. Trial characteristics (eg, blinding, type of intervention and comparator, setting, funding, source of funding, single or multicentric, and postgraduate thesis) were abstracted from each registered trial. A list of all the randomized clinical trials registered on the website was made and their subsequent publication was systematically searched on PubMed and Google Scholar using their registered CTRI number up to December 2021. For trials that were not found, repeat searches were performed, searching by the first author's name, second author's name, and title of the registered trial. The proportion of trials subsequently published and the time to publication from the date of registration were analyzed. Factors associated with publication were compared between trials that were published and those not published using  $\chi^2$  test and univariate analysis by calculating the odds ratio and 95% CI. Multivariable logistic regression was conducted for factors with  $P < .50$ .

**Results** The first 100 pediatric intervention trials registered from 2008 to 2012 were retrieved from the CTRI. The overall proportion of trials published was 71%; 87% had randomization with a comparator arm, 78% examined intervention for treatment, 22% examined intervention for process of care change, 62% had retrospective registration, 51% had funding, and 39% had government funding. The proportion of trials that were performed at a single center was 75%, 92% of trials were hospital based, 84% were conducted at teaching hospitals, and 43% were postgraduate thesis based. The median (range) time to publication was 4 (1-9) years. A non-statistically significant higher proportion of postgraduate thesis-based trials vs non-thesis-based trials (79% vs 65%; OR, 2.04; 95% CI, 0.81-5.09) and single-center trials vs multicenter trials (76% vs 56%; OR, 2.48; 95% CI, 0.96-6.44) were published. On multivariable logistic regression, none of the factors were associated with higher odds of publication (**Table 85**).

**Table 85. Univariate and Multivariable Logistic Regression Analysis of Factors Affecting Publication of Pediatric Intervention Trials**

Factor	No. of trials	Trials published, No. (%)	P value	OR (95% CI)	aOR (95% CI) <sup>a</sup>
Randomization with comparator arm	87	63 (72)	.42	1.64 (0.48 to 5.51)	1.97 (0.53 to 7.26)
Randomization without comparator arm	13	8 (62)			
Blinding	53	37 (70)	.44	0.69 (0.26 to 1.78)	0.65 (0.25 to 1.69)
No blinding	39	30 (77)			
Intervention for treatment	78	55 (71)	.84	0.89 (0.31 to 2.58)	NA
Intervention for process of care change	22	16 (73)			
Not funded	49	37 (76)	.33	0.64 (0.27 to -1.55)	1.24 (0.37 to 4.06)
Funded	51	34 (67)			
Government funding	20	16 (80)	.10	2.88 (0.78 to 18.67)	NA
Private funding	31	18 (58)			
Hospital-based trials	92	66 (72)	.58	1.52 (0.33 to 6.83)	NA
Community-based trials	8	5 (63)			
Postgraduate thesis-based trials	43	34 (79)	.12	2.04 (0.81 to 5.09)	1.57 (0.47 to 5.23)
Non-postgraduate thesis-based trials	57	37 (65)			
Single center	75	57 (76)	.06	2.48 (0.96 to 6.44)	2.38 (0.75 to 7.55)
Multiple center	25	14 (56)			
Teaching hospital	84	60 (72)	.82	1.13 (0.35 to 3.61)	NA
Non-teaching hospital	16	11 (69)			
Retrospective registration	62	45 (73)	.65	1.22 (0.50 to 2.95)	NA
Prospective registration	38	26 (68)			

Abbreviations: aOR, adjusted odds ratio; NA, not applicable; OR, odds ratio.

<sup>a</sup>From multivariable logistic regression.

**Conclusions** Seventy-one percent of pediatric intervention trials registered in the CTRI were subsequently published regardless of their blinding, funding status, type of intervention, comparator, or setting. A non-statistically significant higher proportion of postgraduate thesis-based trials and single-center trials were published. Further studies with a larger sample size are needed to demonstrate any statistical significance. Journal editors, funding agencies, and ethics committees could implement mandatory registration of intervention trials, as registration seems to be associated with better quality and a good chance of subsequent publication. Following guidelines for preparing data sets for submission to data repositories could help achieve more trial registrations.

## References

1. Rao MVV, Maulik M, Juneja A, et al. Clinical Trials Registry—India: a decadal perspective. *Indian J Pharmacol.* 2020;52(4):272-282. doi:10.4103/ijp.IJP\_24\_20
2. DePasse JM, Park S, Eltorai AEM, Daniels AH. Factors predicting publication of spinal cord injury trials registered on www.ClinicalTrials.gov. *J Back Musculoskelet Rehabil.* 2018;31(1):45-48. doi:10.3233/BMR-169628

3. National Heart, Lung, and Blood Institute. Guidelines for preparing clinical study data sets for submission to the NHLBI data repository. Accessed February 20, 2022. <https://www.nhlbi.nih.gov/grants-and-training/policies-and-guidelines/guidelines-for-preparing-clinical-study-data-sets-for-submission-to-the-nhlbi-data-repository>

<sup>1</sup>Department of Pediatrics, University College of Medical Sciences, New Delhi, India, sumairakhalil@yahoo.com; <sup>2</sup>Department of Pediatrics, Maulana Azad Medical College, New Delhi, India

**Conflict of Interest Disclosures** None reported.

## Retractions

### Characterization of Publications on Post-Retraction Citation of Retracted Articles

Jodi Schneider,<sup>1</sup> Randi Proescholdt,<sup>1,2</sup> Jacqueline Leveille,<sup>1</sup> Susmita Das,<sup>1</sup> for the Reducing the Inadvertent Spread of Retracted Science (RISRS) Team

**Objective** Existing literature reviews about retraction do not analyze postretraction citation. This research synthesized a subgroup of empirical studies about retraction and reports what is known about postretraction citation.

**Design** This was a subanalysis of a previously reported scoping review.<sup>1</sup> A total of 386 items about retraction were found by double screening a systematic search of PubMed, Scopus, and Web of Science up to February 10, 2021; Scopus-cited reference searching in January 2021; and hand searching up to July 2021.<sup>1</sup> Items comprised published and unpublished (work in progress) research reports. Subsequent to the previous report, a custom taxonomy<sup>2</sup> of methods, research goals, and research was iteratively created and made searchable via an online bibliography.<sup>3</sup> For the subanalysis, the scope was narrowed to postretraction citation after excluding items on citation-related implications for review literature and authors' careers, publicity, and altmetrics. A codebook to guide data extraction was developed and piloted. Data extraction and analysis is ongoing.

**Results** This subanalysis included 92 items up to July 2021 on postretraction citation of retracted papers. Items were classified into 7 topics: database-focused analyses (n = 33) (eg, PubMed and Web of Science); field-based case studies (n = 20) (eg, genetics, radiology, and thoracic surgery); paper-focused case studies of 1 to 125 selected papers (n = 15); author-focused case studies of 1 or several authors with many retracted publications (n = 15) (eg, Bruening, Darsee, and Reuben); studies of retracted publications cited in review literature (n = 8); geographic case studies (n = 4) (focusing on Brazil, the European Union, Italy, and South Korea); and studies selecting retracted publications by method (n = 2) (eg, human subjects and randomized clinical trials). Five items were classified as belonging to 2 topics each. Empirical research about postretraction citation has been published in a diffuse set of journals, primarily in journals of ethics, information science, meta-science and scientometrics, and

domain sciences, especially medical specialties. The earliest 2 studies identified were both published in 1990; 1 study was an author-focused case study of citations to Stephen E. Breuning's publications and the other studied 82 Index Medicus articles retracted, with retraction notices in Index Medicus as of 1990 and citations from SCISEARCH. From 1990 to 2017, a total of 1 to 3 items were found per year except in 2016 (5 items), and from January 2018 to July 2021, 11 to 18 items per year were found on postretraction citation. Almost all of these items focused on health sciences (eg, medicine, dentistry, nursing, psychology, and pharmacy), with 1 item focused on arts and humanities and 2 items focused on engineering.

**Conclusions** Postretraction citation has been studied consistently since 1990, with increasing attention since 2018. This analysis found an increasing number of items on postretraction citation from January 2018 to July 2021. However, relevant work published after July 2021 was not included, and items published earlier may have been missed from the scoping review. Inclusion of work in progress may have increased publication counts for the most recent years.

## References

1. Schneider J, Woods ND, Proescholdt R, Fu Y; RISRS Team. Recommendations from the Reducing the Inadvertent Spread of Retracted Science: shaping a research and implementation agenda project. *MetaArXiv Preprints*. Preprint posted online July 28, 2021. doi:10.31222/osf.io/ms579
  2. Proescholdt R, Schneider J. Comparison set guide for empirical retraction lit. University of Illinois IDEALS Repository. Deposited September 10, 2021. <https://hdl.handle.net/2142/110389>
  3. Proescholdt R, Schneider J; RISRS Team. Empirical retraction lit bibliography. Version 2.20. Zenodo. September 9, 2021. doi:10.5281/zenodo.5498500
- <sup>a</sup>University of Illinois Urbana-Champaign, Champaign, IL, USA, jodi@illinois.edu; <sup>b</sup>Menlo College, Atherton, CA, USA
- Conflict of Interest Disclosures** Jodi Schneider has been an invited speaker for scholarly publishing organizations Committee on Publication Ethics, CrossRef, the European Association of Science Editors, the International Society of Managing and Technical Editors, the Institute of Electrical and Electronics Engineers, the National Information Standards Organization, and STM; has received data-in-kind from Retraction Watch and scite; and has received usability testing compensation from the Institute of Electrical and Electronics Engineers. No other disclosures were reported.
- Funding/Support** This project was supported by the Alfred P. Sloan Foundation (G-2020-12623).
- Role of the Funder/Sponsor** The study sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the abstract; and decision to submit the abstract for presentation.