Final Program and Abstracts

Tenth International Congress on Peer Review and Scientific Publication

Enhancing the quality and credibility of science

Sept 3-5, 2025 | Swissôtel | Chicago, IL

10th

JAMA Network | the bmj | METRICS META-RESEARCH INNOVATION CENTER AT STANFORD

Tenth International Congress on Peer Review and Scientific Publication

Table of Contents

Welcome 2

Advisory Board 3

Program Highlights 4

Program - Plenary Sessions 5

Program - Poster Sessions 10

Plenary Session Invited Talk Abstracts 19

Plenary Session Abstracts by Topic
Author and Reviewer Use of AI 21
Authorship and Integrity Issues 26
Diversity and the Research Environment 30
Research Misconduct and Integrity 33
Bias, Study Outcomes, and Reporting Concerns 40
Peer Review Models 46
Editorial and Publishing Processes and Models 50
Peer Review Times and Payment Incentives 54
Use of AI to Assess Quality and Reporting 58
Open Science, Availability of Protocols, and Registration 62
Open Science and Data Sharing 68
AI for Detecting Problems and Assessing Quality in Peer Review 70

Poster Session Abstracts by Topic
AI in Peer Review and Publication 75
Authorship and Contributorship 89
Bias 93
Bibliometrics and Publication Metrics 100
Conflict of Interest 105
Data Sharing and Access 114
Diversity and Inclusion 117
Editorial and Peer Review Processes 127
Education/Training 136
Errors and Corrections 140
Funding/Grant Peer Review 142
Misconduct and Research Integrity 147
Open and Public Access 152
Open Science 154
Pandemic Science 158
Paper Mills 160
Peer Review 162
Peer Review Process and Models 167
Predatory Journals 172
Preprints 174
Preregistration of Studies 176
Quality of Reporting 181
Reporting Guidelines 189
Reproducibility 195
Research Methods 199
Retractions 206
Social Media 213

Sponsors 216

Exhibitors 218

Congress Organizers and Planners 221

Tenth International Congress on
Peer Review and Scientific Publication
Enhancing the quality and credibility of science

Welcome

The JAMA Network, The BMJ, and METRICS welcome you to Chicago and the Tenth International Congress on Peer Review and Scientific Publication. Our aim is to encourage research with the quality, integrity, and credibility of peer review and scientific publication, to establish the evidence base on which scientists can improve the conduct, reporting, and dissemination of scientific research. We have continued our efforts to broaden the scope of the Congress to all aspects of peer review and publication—from funding to postpublication—and to all sciences.

We will have 3 days for presentations of new research into peer review and all aspects of scientific publication, bias, quality of reporting, and information access and dissemination. There are 51 plenary session research presentations and 4 plenary invited talks. Each plenary session research presentation will be followed by equal time for discussion and questions from the audience. In addition, in-person poster presentations are scheduled for Thursday and Friday, and additional posters are available online.

This year's meeting is hybrid, and all plenary sessions will be livestreamed and all plenary presentations and posters are available on the meeting platform at via the online platform at underline.io/events/476/reception with opportunities to view presentations, comment and ask questions via chat during and after the meeting.

We hope you will take an active part in the program, as we depend on your participation in the discussion sessions to make the Congress a success. Enjoy the Congress and enjoy Chicago!

Congress Directors
John P.A. Ioannidis
Michael Berkwits

Congress Executive Director
Annette Flanagin

European Director
Theodora Bloom

Follow and tag us on social media!

@peerreviewcongress #PRC10

@peerreviewcongress.bsky.social #PRC10

Peer Review Congress Advisory Board

Program Highlights
Three Days of Original Research

September 3
- Author and Reviewer Use of AI
- Authorship and Integrity Issues
- Diversity and Research Environment
- Research Misconduct and Integrity

September 4
- Bias, Study Outcomes, and Reporting Concerns
- Peer Review Models
- Editorial and Publishing Processes and Models
- Peer Review Times and Payment Incentives

September 5
- Use of AI to Assess Quality and Reporting
- Open Science, Availability of Protocols, and Registration
- Open Science and Data Sharing
- AI for Detecting Problems and Assessing Quality in Peer Review

- 55 Plenary Session Presentations
- 102 In-person Posters
- 40 Virtual Posters

Equal time for presentation and audience participation

Plenary Session Invited Talks

The Drummond Rennie Lecture

Forward to the Past—Making Contributors Accountable

Ana Marušić (Croatia)

The Douglas G. Altman Lecture

Does the Journal Article Have a Future?

Malcolm MacLeod (Scotland)

Journal Prestige Can and Should Be Earned

Simine Vazire (Australia)

A Singular Disruption of Scientific Publishing—AI Proliferation and Blurred Responsibilities of Authors, Reviewers, and Editors

Zak Kohane (US)

Land Acknowledgement

The City of Chicago is located on land that is and has long been a center for Native peoples. The area is the traditional homelands of the Anishinaabe, or the Council of the Three Fires: the Ojibwe, Odawa, and Potawatomi Nations. Many other Nations consider this area their traditional homeland, including the Myaamia, Ho-Chunk, Menominee, Sac and Fox, Peoria, Kaskaskia, Wea, Kickapoo, and Mascouten. The City specifically acknowledges the contributions of Kitihawa of the Potawatomi in fostering the community that has become Chicago. We acknowledge all Native peoples who came before us and who continue to contribute to our City. We are committed to promoting Native cultural heritage.

Meeting Code of Conduct

The Peer Review Congress follows a Code of Conduct to ensure there is a professional and ethical environment for all attendees. Everyone should feel welcome, safe, and able to participate without fear of unwelcome conduct in-person or through electronic communication. Attendees should be respectful of others, identify themselves by name and affiliation when speaking, and declare conflicts of interest. The Code of Conduct includes physical, verbal, and nonverbal behavior and applies to in-person and virtual/remote participation. If you have any questions about the Code of Conduct, are aware of behavior that may have violated this policy, or would like to report an incident, you can contact the Code of Conduct liaison for the Peer Review Congress Andrew Roff at andrew.roff@ama-assn.org.

Program

All plenary sessions will be held in the Zurich Ballroom, D-F.
Poster sessions will be held in St Gallen and Montreux rooms.
Breaks and Exhibits will be held in Zurich Foyer and Zurich A-C.
Luncheons will be served in the Vevey room.

Plenary Sessions

Wednesday, September 3

7:00 AM - 8:00 AM

Registration, Breakfast, and Visit Exhibits

8:00

Welcome
John Ioannidis (United States)

8:05 AM - 8:30 AM

The Drummond Rennie Lecture
Forward to the Past—Making Contributors Accountable
Ana Marušić (Croatia)

8:30 AM - 9:50 AM

Author and Reviewer Use of AI
Moderator: Christine Laine (United States)

Authors Self-disclosed Use of Artificial Intelligence in Research: Submissions to 49 BMJ Group Biomedical Journals
Isamme AlFayyad, Maurice Zeegers, Lex Bouter, Helen Macdonald, Sara Schroter (Netherlands, Saudi Arabia, United Kingdom)

Artificial Intelligence Use and Acknowledgment in Medical Research Writing Among Chinese Scholars
Yang Zhang, Mengyuan Duan, Guoguang Zhao, Penghu Wei, Xiuyuan Hao (China)

Factors Associated With Author and Reviewer Declared Use of AI in Medical Journals
Roy Perlis, Annette Flanagin, Michael Berkwits, Jacob Kendall-Taylor, Kirsten Bibbins-Domingo (United States)

Quantifying and Assessing the Use of Generative AI by Authors and Reviewers in the Cancer Research Field
Daniel Evanko, Michael Di Natale (United States)

9:50 AM - 10:20 AM

Refreshment Break and Visit Exhibits

10:20 AM - 12:00 PM

Authorship and Integrity Issues
Moderator: Lex Bouter (Netherlands)

Comparison of Reasons for Retraction of Biomedical Articles by Women and Men Authors
Ana-Catarina Pinho-Gomes, Carinna Hockham, Mark Woodward (United Kingdom)

Paper Mill Use of Fake Personas to Manipulate the Peer Review Process
Tim Kersjes (Netherlands)

Authorship Changes as an Indicator of Research Integrity Concerns in Submissions to Academic Journals
Coromoto Power Febres, Julia Gunn, Laura Wilson (United Kingdom, United States)

Notifying Authors That They Have Cited a Retracted Article and Future Citations of Retracted Articles: The RetractoBot Randomized Controlled Trial
Nicholas DeVito, Christine Cunningham, Ben Goldacre (United Kingdom)

How a Questionable Research Network Manipulated Scholarly Publishing
Leslie McIntosh, Hélène Draux, Elizabeth Smee, Cynthia Hudson Vitale (United Kingdom, United States)

12:00 PM - 1:30 PM

Lunch and Visit Exhibits

1:30 PM - 2:30 PM

Diversity and Research Environment
Moderator: Vivienne Bachelet (Chile)

An Analysis of Equity, Diversity, and Inclusion Concerns From JAMA Network Peer Reviewers
Michael Mensah, Anand Habib, Jacob Kendall-Taylor, Mya Roberson, Kanade Shinkai, Annette Flanagin, Preeti Malani (United States)

Assessment of an Intervention to Equalize the Proportion of Funded Grant Applications for Underrepresented Groups at the Canadian Institutes of Health Research
Anne Lasinsky, James Wrightson, Matthew Hogel, Alannah Brown, Adrian Mota, Karim Khan, Clare Ardern (Canada)

Extracting Research Environment Indicators From the UK Research Excellence Framework 2021 Statements
Noémie Aubert Bonn, Lukas Hughes-Noehrer (United Kingdom)

2:30 PM - 3:10 PM

Refreshment Break and Visit Exhibits

3:10 PM - 5:30 PM

Research Misconduct and Integrity
Moderator: David Schriger (United States)

Characterizing Problematic Images in Retracted Scientific Articles
João Phillipe Cardenuto, Daniel Moreira, Anderson Rocha (Brazil, United States)

Misidentification of Scanning Electron Microscope Instruments in the Peer-Reviewed Materials Science and Engineering Literature
Reese Richardson, Jeonghyun Moon, Spencer Hong, Luís Amaral (United States)

Retractions and Democracy Index Scores Across 167 countries
Ahmad Sofi-Mahmudi, Hesam Salmabadi (Canada)

Indicators of Small-Scale and Large-Scale Citation Concentration Patterns
Iakovos Evdaimon, John Ioannidis, Giannis Nikolentzos, Michail Chatzianastasis, George Panagopoulos, Michalis Vazirgiannis (France, Greece, Luxembourg, United States)

Scale and Resilience in Organizations Enabling Systematic Scientific Fraud
Reese Richardson, Spencer Hong, Jennifer Byrne, Thomas Stoeger, Luís Amaral (Australia, United States)

Patterns of Paper Mill Papers and Retraction Challenges
Anna Abalkina, Svetlana Kleiner (Germany, Netherlands)

Sustainable Approaches to Upholding High Integrity Standards in the Face of Large-Scale Threats: Insights From *PLOS One*
Renee Hoch, Emily Chenette (United Kingdom, United States)

5:30 PM - 6:30 PM

Welcome Reception

Thursday, September 4

8:00 AM

Welcome
Mike Berkwits (United States)

8:05 AM

The Douglas G. Altman Lecture
Does the Journal Article Have a Future?
Malcolm MacLeod (Scotland)

8:30 AM - 10:10 AM

Bias, Study Outcomes, and Reporting Concerns
Moderator: Steve Goodman (United States)

Immortal Time Bias Prevalence and Effects on Estimates in Systematic Reviews and Meta-Analyses
Jae Il Shin, Minseo Kim, Dong Keon Yon, Seung Won Lee, Masoud Rahmati, Marco Solmi, André Carvalho, Ai Koyanagi, Lee Smith, John Ioannidis (Australia, Canada, Iran, South Korea, Spain, United Kingdom, United States)

Effect Estimates for the Same Outcomes Designated as Primary vs Secondary in Randomized Clinical Trials: A Meta-Research Study
Yiwen Jiang, Yuanxi Jia, Karen Robinson, Jinling Tang (China, Singapore, United States)

Detection and Monitoring of Potential Outcome Reporting Bias Using Large Language Models: Application to FDA-Regulated Drug Trials
Ian Bulovic, Susmitha Wunnava, Wonjin Yoon, Adam Dunn, Timothy Miller, Florence Bourgeois (Australia, United States)

Data Repurpose in AI Studies and Scientific Outcomes
Yulin Yu, Yong-Yeol Ahn, Daniel Romero (United States)

Prevalence of the Statement "to Our Knowledge" and Similar Paraphrases in Current and Past Biomedical Literature
Nicola Di Girolamo, Reint Meursinge Reynders, Ugo Di Girolamo (Netherlands, United States)

10:10 AM - 10:40 AM

Refreshment Break and Visit Exhibits

10:40 AM - 12:00 PM

Peer Review Models

Moderator: Véronique Kiermer (United States)

Peer Reviews of Peer Reviews: A Randomized Controlled Trial and Other Assessments
Alexander Goldberg, Ivan Stelmakh, Kyunghyun Cho, Alice Oh, Alekh Agarwal, Danielle Belgrave, Nihar Shah (Russia, South Korea, United Kingdom, United States)

Anonymizing Reviewers to Each Other in Peer Review Discussions: A Randomized Controlled Trial
Charvi Rastogi, Xiangchen Song, Zhijing Jin, Ivan Stelmakh, Hal Daume, Kun Zhang, Nihar Shah (Russia, United States)

Dual Anonymous and Distributed Peer Review for Proposal Review Rankings at the ALMA Observatory
John Carpenter, Andrea Corvillon (Chile)

Comparison of Content in Published and Unpublished Peer Review Reports
Elena Álvarez- García, Daniel Garcia-Costa, Flaminio Squazzoni, Mario Malički, Bahar Mehmani, Francisco Grimaldo (Italy, Netherlands, Spain, United States)

12:00 PM - 1:30 PM

Lunch and Visit Exhibits

1:30 PM - 2:30 PM

Editorial and Publishing Processes and Models

Moderator: James Kigera (Kenya)

Changes to Research Article Abstracts Between Submission and Publication
Christos Kotanidis, Sarah Gorey, Harleen Marwah, Abarna Pearl, Darren Taichman, Mary Beth Hamel (United States)

Manuscript Characteristics Associated With Editorial Review and Peer Review Outcomes at *Science* and *Science Advances*
Nicholas LaBerge, Sam Zhang, Daniel Larremore, Aaron Clauset (United States)

Investigating Changes in Common Vocabulary Terms in *eLife* Assessments Across Versions in a Publish, Review, Curate Model
Nicola Adamson, Andy Collings (United Kingdom)

2:30 PM - 3:45 PM

Poster Sessions, Refreshment Break, and Visit Exhibits

3:45 PM - 5:05 PM

Peer Review Times and Payment Incentives

Moderator: Kirsten Bibbins-Domingo (United States)

Results of Testing the Gold Standard 2-Week Reviewer Deadline
Emilie Gunn, Kelly Brooks, Stephanie Valladao (United States)

Analysis of Decisions and Lead-Time in Ethical Review Boards in Sweden
Emmanuel Zavalis, Love Ahnström, Natasha Ohlsson, Gustave Nilsonne (Sweden)

Monetary Incentives for Peer Review at a Medical Journal: A Quasi-Randomized Experimental Study
Christopher Cotton, Abid Alam, Sophie Tosta, Timothy Buchman, David Maslove (Canada, United States)

Exploring Views on Remuneration for Review: A Survey of *BMJ*'s Patient and Public Reviewers
Sara Schroter, Rebecca Harmstron, Emma Doble, Sophie Cook, Amy Price (United Kingdom, United States)

5:05 PM - 5:35 PM

Journal Prestige Can and Should Be Earned
Simine Vazire (Australia)

Friday, September 5

8:00 AM

Welcome
Adrian Aldcroft (United Kingdom)

8:05 AM

A Singular Disruption of Scientific Publishing—AI Proliferation and Blurred Responsibilities of Authors, Reviewers, and Editors
Zak Kohane (US)

8:30 AM - 9:50 AM

Use of AI to Assess Quality and Reporting
Moderator: Timothy Feeney (United States)

Natural Language Processing to Assess the Role of Preprints in COVID-19 Policy Guidance
Nicholas Evans, Samuel Angelli-Nichols, Emma Chang-Rabley, Yara Omar, Rachel Nas, Mikaela Finnegan, Rocco Casagrande, Emily Ricotta (United States)

Leveraging Large Language Models for Assessing the Adherence of Randomized Controlled Trial Publications to Reporting Guidelines
Lan Jiang, Xiangji Ying, Mengfei Lan, Andrew Brown, Colby Vorland, Evan Mayo-Wilson, Halil Kilicoglu (United States)

Understanding How a Language Model Assesses the Quality of Randomized Controlled Trials: Applying Shapley Additive Explanations to Encoder Transformer Classification Models
Fangwen Zhou, Muhammad Afzal, Rick Parrish, Ashirbani Saha, R. Brian Haynes, Alfonso Iorio, Cynthia Lokker (Canada, United Kingdom)

Using GPT to Identify Changes in Clinical Trial Outcomes Registered on ClinicalTrials.gov
Xiangji Ying, Colby Vorland, Kiran Ninan, Jean-Pierre Oberste, Andrew Brown, Riaz Qureshi, Sirui Zhang, Nicholas DeVito, Matthew Page, Ian Saldanha, Halil Kilicoglu, Evan Mayo-Wilson (Australia, United Kingdom, United States)

9:50 AM - 10:20 AM

Refreshment Break and Visit Exhibits

10:20 AM - 12:00 PM

Open Science, Availability of Protocols, and Registration
Moderator: Isabelle Boutron (France)

Perceived Risks and Barriers to Open Research Practices in UK Higher Education
Lukas Hughes-Noehrer, Noémie Aubert Bonn (United Kingdom)

Use of an Open Science Checklist and Reproducibility of Findings: A Randomized Controlled Trial
Ayu Putu Madri Dewi, Melissa Rethlefsen, Sara Schroter, Florian Naudet, Nicholas DeVito, Constant Vinatier, Inge Stegeman, Mariska Leeflang, Gowri Gopalakrishna (France, Netherlands, United Kingdom, United States)

Nonregistration, Discontinuation, and Nonpublication of Randomized Trials in Switzerland, the UK, Germany, and Canada: An Updated Meta-Research Study
Benjamin Speich, Ala Heravi, Johannes Schwenke, Christof Schönenberger, Lena Hausheer, Dmitry Gryaznov, Jason Busse, Manuela Covino, Szimonetta Lohner, Malena Chiaborelli, Ruben Ramirez, Ramon Saccilotto, Erik von Elm, Arnav Agarwal, Julian Hirt, David Mall, Alain Amstutz, Selina Epp, Dominik Mertz, Anette Blümle, Belinda von Niederhäusern, Ayodele Odutayo, Alexandra Griessbach, Sally Hopewell, Matthias Briel (Canada, Germany, Hungary, Switzerland, United Kingdom)

Factors Associated With Improper Clinical Trial Registration, Registration Deficiencies, and Publication Status of Submissions to *The BMJ*
David Blanco, Elizabeth Loder, Sophie Cook, Sara Schroter (Spain, United Kingdom, United States)

Registered Clinical Trial Trends in East Asia and the United States, 2014 to 2025
Eunhye Lee, San Lee, Jae Il Shin, John Ioannidis (South Korea, United States)

12:00 PM - 1:30 PM

Lunch and Visit Exhibits

1:30 PM - 2:30 PM

Open Science and Data Sharing

Moderator: Valda Vinson (United States)

Researcher Adherence to Journal Data Sharing Policies: A Meta-Research Study

Aidan Tan, Yiyi Lin, Michellie Lian, Zhilin Ren, Tony Lian, Vincent Yuan, Angela Webster, Anna Seidler (Australia)

A Funder-Led Intervention to Increase the Sharing of Data, Code, Protocols, and Key Laboratory Materials

Robert Thibault, Dana Cobb-Lewis, Matt Lewis, Devin Snyder, Cornelis Blauwendraat, Sonya Dumanis (United States)

Medical Journal Policies on Requirements for Clinical Trial Registration, Reporting Guidelines, and Data Sharing: A Systematic Review

Kyobin Hwang, Z exing Song, Marsida Stafa, Jodie Chiu, An-Wen Chan (Canada)

2:30 PM - 3:45 PM

Poster Sessions, Refreshment Break, and Visit Exhibits

3:45 PM - 5:05 PM

AI for Detecting Problems and Assessing Quality in Peer Review

Moderator: John Ioannidis (United States)

Leveraging Large Language Models for Detecting Citation Quotation Errors in Medical Literature

M. Janina Sarol, Jodi Schneider, Halil Kilicoglu (United States)

Automating the Detection of Promotional (Hype) Language in Biomedical Research

Bojan Batalo, Neil Millar, Erica Shimomoto (Japan)

Evaluation of a Method to Detect Peer Reviews Generated by Large Language Models

Vishisht Rao, Aounon Kumar, Himabindu Lakkaraju, Nihar Shah United States)

Quality and Comprehensiveness of Peer Reviews of Journal Submissions Produced by Large Language Models vs Humans

Fares Alahdab, Juan Franco, Helen Macdonald, Sara Schroter (United States, United Kingdom)

5:05 PM

Closing

Poster Session Abstracts

All In-person Posters will be presented on Thursday, September 4, and Friday, September 5; In-person and Virtual Posters will be available to view and post comments and questions via the online meeting platform at underline.io/events/476/reception All posters and related materials will be available online after the meeting.

AI in Peer Review and Publication

Domain-Specific Pretrained Encoder Transformers for the Identification of Methodologically Rigorous Systematic Reviews: A Retrospective Modeling Study

Fangwen Zhou, Muhammad Afzal, Rick Parrish, Ashirbani Saha, Wael Abdelkader, R. Haynes, Alfonso Iorio, Cynthia Lokker (Canada, United Kingdom)

Strategic Insights Into Editor Engagement With AI-Assisted Tools Based on Survey and Data Analysis of AI-Assisted Ethics Checks

Beth Waymouth, Heather Slater, Angharad Goode, Katie Allin, Maria Kowalczuk (Switzerland, United Kingdom)

Attitudes and Perceptions of Biomedical Journal Editors in Chief Toward the Use of Artificial Intelligence Chatbots in the Scholarly Publishing Process

Jeremy Ng, Malvika Krishnamurthy, Gursimran Deol, Wid Al-Khafaji, Vetrivel Balaji, Magdalene Abebe, Jyot Adhvaryu, Tejas Karrthik, Pranavee Mohanakanthan, Adharva Vellaparambil, Lex Bouter, R. Haynes, Alfonso Iorio, Cynthia Lokker, Hervé Maisonneuve, Ana Marušić, David Moher (Canada, Croatia, France, Germany, Netherlands)

Attitudes and Perceptions Toward the Use of Artificial Intelligence Chatbots in Medical Journal Peer Review: A Large-Scale, International Cross-Sectional Survey

Jeremy Ng, Daivat Bhavsar, Neha Dhanvanthry, Lex Bouter, Teresa Chan, Holger Cramer, Annette Flanagin, Alfonso Iorio, Cynthia Lokker, Hervé Maisonneuve, Ana Marušić, David Moher (Canada, Croatia, France, Germany, Netherlands, United States)

Usefulness of LLMs as an Author Checklist Assistant for Scientific Papers: NeurIPS'24 Experiment

Alexander Goldberg, Ihsan Ullah, Thanh Gia Hieu Khuong, Benedictus Rachmat, Zhen Xu, Isabelle Guyon, Nihar Shah (France, United States)

Accuracy and Precision of a Neural Network Author Name Disambiguator

Vicente Amado Olivo, Wolfgang Kerzendorf, Nutan Chen, Joshua Shields, Bangjing Lu, Andreas Flörs (Germany, United States)

An AI-assisted Analysis of Published *PeerJ* Open Peer Reviews

Peiling Wang, Dietmar Wolfram, Scott Shumate (United States)

Comparing Observational Exposure-Phenotype Correlations With Large Language Model Predictions

Chirag Patel, Arjun Manrai, Randall Ellis, John Ioannidis (United States)

Pragmatic Assessment of Different AI Large Language Models for Extraction of CONSORT Items From Randomized Controlled Trials Before Peer Review

Nicola Di Girolamo, Reint Meursinge Reynders, Ugo Di Girolamo (Netherlands, United States)

Artificial Intelligence Editorial Policies and Reporting Standards in Orthopedic and Sports Medicine Journals

Josh Major, Kurt Mahnken, Alec Young, Cameron O'Brien, Andrew Tran, Patrick Crotty, Alica Ford, Matt Vassar (United States)

Enhancing Research Integrity in Abstract Submissions With a Hybrid AI-Human Post-Review Process

Heather Goodell, Chirag (Jay) Patel Jonathan Schultz, Christine Beaty, Shilpi Mehra (United Arab Emirates, United States)

Virtual

Policies on Artificial Intelligence Among Academic Publishers

Jeremy Ng, Daivat Bhavsar, Laura Duffy, Hamin Jo, Cynthia Lokker, R.Brian Haynes, Alfonso Iorio, Ana Marušić (Canada, Croatia, Germany)

Use of Generative Artificial Intelligence Tools by Authors and Reviewers of *Eurosurveillance*
Eva Sarachaga, Ines Steffens (Sweden)

Use of an AI Peer Review Panel to Assess Clarity, Novelty, and Impact
Pawin Taechoyotin, Daniel Acuna (United States)

Reviewer Rating Variability and Confidence and Language Model Sentiment Prediction of Machine Learning Conference Papers
Yidan Sun, Mayank Kejriwal (United States)

Authorship and Contributorship

Co–First Authors and Co–Corresponding Authors in the *Chinese Medical Journal* and *JAMA*
Ting Gao, Xiuyuan Hao (China)

Authorship and Contributorship Criteria and Practices at the *Annals of African Surgery*
Cecilia Munguti, James Kiilu, James Kigera, Michael Mwachiro (Kenya)

Integration of Credit and Accountability Principles in Authorship Policies of Science Journals and Research Institutions
Mohammad Hosseini, Sofie Adams, Yensi Flores, Kathleen Jamieson, Joerg Heber, Jennifer Heimberg, Véronique Kiermer, Arthur Lupia, Ana Marušić, Beau Nielsen, Magdalena Skipper, Geeta Swamy, Susan Wolf (Croatia, Ireland, United Kingdom, United States)

Authors Who Publish in a Journal and Likelihood to Serve as Reviewers
Stephan Fihn, Roy Perlis Jacob Kendall-Taylor, Annette Flanagin (United States)

Virtual

Compiling the Publications Produced by Medical Writing
Maud Bernisson (France)

Bias

Influence of Promotional Language on Evaluations of Biomedical Literature: A Randomized Controlled Trial
Brian Budgell, Neil Millar (Canada, Japan)

Unraveling the Spin and Selective Reporting in Medical AI Research: A Cross-Sectional Meta-Research Study
Vincent Yuan, Aidan Christopher Tan (Australia)

A Comparison of Self-Acknowledged Limitations With Risk of Bias Assessments
Joe Menke, Mengfei Lan, Halil Kilicoglu (United States)

Study Hypotheses and Results From Superiority and Noninferiority Randomized Clinical Trials
Yuanxi Jia, Yiwen Jiang, Karen Robinson, Jinling Tang (China, Singapore, United States)

Assessment of Spin Among Diagnostic Accuracy Meta-Analyses Published in Top Pathology Journals: A Systematic Review
Griffin Hughes, Andrew Tran, Sydney Marouk, Eli Paul, Matt Vassar (United States)

Virtual

Topic and Knowledge-Base Interdisciplinarity in Manuscripts Submitted to Physical Science Journals vs Editorial Decision and Reviewer Positivity
Sidney Xiang, Daniel Romero, Misha Teplitskiy (United States)

Biomedical Studies Published With Negative Results Over the Past Decade
Hannah Varkey, Florian Thomas, Elli Gourna Paleoudis (United States)

Bibliometrics and Publication Metrics

Review and Publication Times Across Journals Publishing on Health Policy
Kathryn Phillips, Danea Horn (United States)

Funding Sources and the Online and Academic Impact of Cardiovascular Trials Published in Highest-Impact Journals
Farbod Zahedi Tajrishi, Sina Rashedi, Ashkan Hashemi, Isaac Dreyfus, Nicholas Varunok, John Burton, Björn Redfors, Gregory Piazza, Joshua Wallach, Lesley Curtis, Sanjay Kaul, David Cohen, Roxana Mehran, Flavia Geraldes, Joseph Ross, Jane Leopold, Harlan Krumholz, Gregg Stone, Behnood Bikdeli (United Kingdom, United States)

Virtual

Publication Trends on Priority Epidemics According to the Sustainable Development Goals in Pharmaceutical Journals, 2000-2024
Julia Soto Rizzato, Marcus Silva, David Moher, Tais Galvao (Brazil, Canada)

Science Journal Abstracts Misregistered in the Crossref Database
Qinyue Liu, Yagmur Ozturk, Cyril Labbé (France)

Trends in Citation Impacts of Original Research in Major Cardiovascular Journals, 2008-2018

Younwoo Ki, Chungsoo Kim, Yuan Lu, Joshua Wallach, Behnood Bikdeli, Milton Packer, Harlan Krumholz, Seng Chan You (South Korea, United States)

Conflict of Interest

Psychiatry Editors in Chief Publishing Practices in Their Own Journals

Justin Nguyen, Robert Rubin (United States)

Prevalence and Nature of Conflict of Interest Disclosures in Published Health Technology Assessment Reports

Miro Vukovic, Ana Marušić (Croatia)

Development of a Tool for Addressing Conflicts of Interest in Trials (TACIT) for Use in Systematic Reviews

Andreas Lundh, Isabelle Boutron, Lesley Stewart, Asbjørn Hróbjartsson (Denmark, France, United Kingdom)

Corporate Influence on Peer-Reviewed Research: Insights From BP's Deepwater Horizon Response

Marc-André Gagnon, Blue Miaoran Dong (Canada)

Conflicts of Interest in Research Across Scholarly Disciplines

Helena Van Beersel Krejcikova, Christoffer Korfitsen, Lisa Bero, Jason Dana, David Dorman, Quinn Grundy, Ibo van de Poel, Morten Rosenmeier, Asbjørn Hróbjartsson, Andreas Lundh (Canada, Denmark, Netherlands, United States)

A Taxonomy-Based Guideline Framework for Conflict of Interest Disclosures (CoST)

Pritha Sarkar, Ruth Whittam, Leslie McIntosh (United Kingdom, United States)

Navigating the Challenges of Competing Interest Disclosures in Academic Publishing

Julia Gunn, Coromoto Power Febres, Laura Wilson (United Kingdom, United States)

Virtual

Conflict of Interest Network Robustness and Funder Homogeneity Associate With Adverse Events and Deaths

S. Scott Graham, Joshua Barbour, Zoltan Majdik, Madeline Bruegger, Carlee Baker, Justin Rousseau (United States)

Disclosed and Undisclosed Conflicts of Interest in US Guidelines for the Management of Obesity

Alessandro Bianconi, Matteo Fiore, Maria Flacco, Lamberto Manzoli (Italy)

Data Sharing and Access

Data Availability Statements in Health Research in Articles and Journal Policies in Korea

Sue Kim, Soo Young Kim, Hyun Jung Yi (South Korea)

Metrics of Primary and Secondary Publications of Clinical Trials With Data Shared on the YODA Project Platform

Erfan Taherifard, Hollin Hakimian, Maryam Mooghali, Sahil Mane, Mengyuan Fu, Stephen Bamford, Karla Childers, Nihar Desai, Cary Gross, Debbie Hewens, Harlan Krumholz, Richard Lehman, Jessica Ritchie, Tamsin Sargood, Joshua Wallach, Molly Willeford, Joseph Ross (United Kingdom, United States)

Data Sharing Statement Reporting Across Medical Specialties

Eli Paul, Griffin Hughes, Alex Hagood, Matt Vassar (United States)

Diversity and Inclusion

Retraction Prevalence and Gender Imbalance Among Highly-Cited Authors and Among All Authors Across Scientific Disciplines

John Ioannidis, Angelo Maria Pezzullo, Antonio Cristiano, Guillaume Roberge, Stefania Boccia, Jeroen Baas (Canada, Italy, Netherlands, United States)

Academic Institutional Affiliations and Gender of Authors, Editorial Board Members, and Editors of Journals

Ulrike Muller, Marie Schwaner, Ksenia Keplinger (Germany, United States)

Diversity Among Reviewers Assigned to Evaluate a Paper as a Factor in Diversifying Perspective and Improving the Peer Review Process in Computer Science

Navita Goyal, Ivan Stelmakh, Nihar Shah, Hal Daumé III (Russia, United States)

Geographical Representation of Author Country Among Peer Reviewers and Publishing Success at 60 STEM Journals

James Zumel Dumlao, Misha Teplitskiy (United States)

Author Responses to Editorial Guidance on Reporting of Sex, Gender, Race and Ethnicity Data

Mabel Chew, Taissa Vila, Jashelle Caga-Meller, Zoe Mullan, Diana Samuel (Australia, Brazil, United Kingdom)

Representation of Authors from Low- and Middle-Income Countries (LMICs) in Trials With Participants From LMICs

Harleen Marwah, Abarna Pearl, Christos Kotanidis, Sarah Gorey, Darren Taichman, Mary Beth Hamel (United States)

Evolution of Authorship Diversity in African Surgical Research Over 2 Decades

Vincent Kipkorir, Godfrey Philipo, Mumba Chalwe-Kaja, Tihitena Negussie, Michael Mwachiro, Robert Parker, Seke Kazuma, Stella Itungu, Abebe Bekele (Ethiopia, Kenya, Tanzania, Uganda, Zambia)

Virtual

Integrating Indigenous Knowledge Into Peer Review Processes in Nigerian Environmental and Health Research

Oludele Solaja (Nigeria)

Geographical Disparities in Navigating Rejection in Scientific Publication

Hong Chen, Chris Rider, David Jurgens, Misha Teplitskiy (United States)

Editorial Landscape of Journals in Kenya, Ethiopia, Nigeria, and Mozambique

Patrick Amboka, Daniel Krugman, Tony Aloo (Kenya, United States)

Editorial and Peer Review Processes

Effects of Peer Review and Editorial Workflows in Decision-Making at a Diamond Open Access Journal

Tais Galvao, Everton Silva, Jorge Barreto, Marcus Silva (Brazil)

Reminding Peer Reviewers to Comment on Reporting Items as Instructed by the Journal: An Analysis of 2 Randomized Trials

Hillary Wnfried Ramirez, Malena Chiaborelli, Christof Schönenberger, Katie Mellor, Alexandra Griessbach, Paula Dhiman, Pooja Gandhi, Szimonetta Lohner, Arnav Agarwal, Ayodele Odutayo, Michael Schlussel, Philippe Ravaud, David Moher, Matthias Briel, Isabelle Boutron, Sally Hopewell, Sara Schroter, Benjamin Speich (Canada, France, Hungary, Switzerland, United Kingdom)

Differences Between Manuscripts Versions: A Living Review and Series of Meta-Analyses

Mario Malički, Ana Jerončić, Gerben Ter Riet, Lex Bouter, John Ioannidis, IJsbrand Aalbersberg, Steven Goodman (Croatia, Netherlands, United States)

Quality of Patient Reviewer Comments and Association With Author and Editor Responses

Melecia Miller, Vera Nezgovorova, Ilana Kersch, Mohamed Elsaid, Marina Broitman (United States)

Factors Associated With Outcomes of Appeals of Manuscripts Initially Rejected by a General Medical Journal

Matthew Stanbrook, Shannon Charlebois, George Tomlinson, Meredith Weinhold (Canada)

Automated Targeted Emails for Improving Author Compliance With Study Reporting Requirements and Other Editorial Processes

Daniel Evanko, Deondre Jordan (United States)

Virtual

Trends in Peer Review Metrics at the *Annals of African Surgery*

James Kiilu, Cecilia Munguti, James Kigera, Michael Mwachiro (Kenya)

Impact of a Novel Checklist on the Peer Review Process

Jorge Finke, Sumi Sexton (United States)

Identifying Methodological Concerns in Agency for Healthcare Research and Quality Evidence-Based Practice Center Reports: Analysis of Editorial Review Comments

Haley Holmer, Edi Kuhn, Camber Hansen-Karr, Ed Reid, Mark Helfand (United States)

Education/Training

Peer Review Exercises to Enhance Trainees' Readiness to Confront Unfair or Biased Reviews

Franki Kung, Mariam Aly, Shahana Ansari, Eliana Colunga, M. J. Crockett, Amanda Diekman, Pablo Gomez, Paul McKee, Miriam Pérez, Sarah Stilwell, Matthew Goldrick (United States)

Virtual

A Pilot Program for Early Career Mentorship in Journal Peer-Review

Anjali Garg, Preeti Panda, Lydia Furman, Kimberly Montez, Alex Kemper, Lewis First (United States)

In-Person Peer Review Training to Improve Preparedness to Evaluate Manuscripts

Marcus Silva, Tais Galvao (Brazil)

Quality of an Educational Program to Empower Early Career Faculty and Trainees Through Mentored Training in Peer Review

Susan Galandiuk, Vaitheesh Jaganathan, Hillary Simon (United States)

Errors and Corrections

Quotation Inaccuracy in Medicine: A Systematic Review and Meta-Analysis

Christopher Baethge, Hannah Jergas (Germany)

Virtual

Taiwanese Researchers' Perceptions of Errors and Their Coping Strategies

Chien Chou (Taiwan)

Funding/Grant Peer Review

Multistakeholder Perspectives on Current Attitudes Toward Unmasking Reviewers' Identity in Biomedical Research Proposals' Peer Review: A Qualitative Study

Seba Qussini, Farizah Anami, Kris Dierickx (Belgium, Qatar)

Influence of Using a Systematic Review to Justify New Research on Funding Application Score

Jong-Wook Ban, Hans Lund, Karen Robinson, Ida Svege, Jan-Ole Hesselberg (Norway, United States)

Virtual

Construction and Validation of Instruments for the Peer Review of Grant Proposals in Peru

Max Carlos Ramírez-Soto, Laura Alvarado-Barbarán, Dianeth Rojas-Naccha, Ayda Luna-Mercado, Arlet Arce-Zavala (Peru)

Reviewers' Interpretation and Application of Research Quality Criteria in Grant Peer Review

Rachel Claus (Canada)

Experiences and Challenges Faced by Canadian Health Research Grant Peer Reviewers

Joanie Sims Gould, Anne Lasinsky, Adrian Mota, Karim Khan, Clare Ardern (Canada)

Misconduct and Research Integrity

Experience With 12 Years of Plagiarism and Duplication Screening

Markus Heinemann, Andreas Boening, Kazunori Okabe, Jessica Bogensberger (Japan, Germany)

Vulnerability of Automated Text-Matching-Based Reviewer Assignments to Collusions

Jhih-Yi Hsieh, Aditi Raghunathan, Nihar Shah (United States)

Identifying Potential Duplicate Publications in the Scientific Literature Using Crossref

Cyril Labbé, Qinyue Liu, Amira Barhoumi, Olessya Miroshnichenko (France)

Tortured Phrases as a Sign of Possible Misconduct in Proceedings From an Engineering Conference

Wendeline Swart, Ophélie Fraisier-Vannier, Guillaume Cabanac (France)

Virtual

Plagiarism and Publication Frauds Revealed by Dissernet

Larisa Melikhova, Andrey Rostovtsev, Vasiliy Vlassov (Israel, Russia)

Citation Biases and Citation-Boosting Strategies: A Scoping Review of Predictors

Hans Lund, Karen Robinson, Jong-Wook Ban, Karen Lie, Birgitte Nørgaard (Denmark, Norway, United States)

Open and Public Access

Virtual

Assessing the Cost-Effectiveness of Open Access Publishing in Dermatology Journals

Dante Dahabreh, Kenny Thien Long Ta, Angela Loczi-Storm, Olivia Lim, Dana Chen, Tasneem IS, Alexandria Kristensen-Cabrera, Rahib Islam, Robert Dellavalle, Eamonn Maher (United States)

Public Access to Information Cited in Rare Disease Reports

Mengyuan Fu, Kexin Ling, Xinyi Zhou, Sneha Dave, Can Li, Luwen Shi, Xiaodong Guan, Joseph Ross (China, United States)

Open Science

Citations of Articles With Open Science Indicators in the French Open Science Monitor Dataset

Giovanni Colavizza, Lauren Cadwallader, Iain Hrynaszkiewicz (Italy, United Kingdom)

Automated Interpretation of Statistical Tables in Economics: Prevalence of Reporting Errors and Effectiveness of Open Science Policies

Stephan Bruns, Helmut Herwartz, John Ioannidis, Chris Islam, Fabian Raters (Belgium, Germany, United States)

Detection of Open Science Practices in Major Medical Journals: A Survey and Diagnostic Accuracy of Automatic Tools Using Sensitivity and Specificity

Constant Vinatier, Ayu Putu Madri Dewi, Gwénaël Dumont, Tracey Weissgerber, Vladislav Nachev, Gowri Gopalakrishna, Maud Scheidecker, François-Joseph Arnault, Nicholas DeVito, Guillaume Freyermuth, Mathieu Acher, Gauthier Le Bartz Lyan, Inge Stegeman, Mariska Leeflang, Florian Naudet (France, Germany, Netherlands, Portugal, United Kingdom)

Pandemic Science

Analysis of Editorials on the Response to the H1N1 and COVID-19 Pandemics

Luka Ursic, Nensi Bralić, Giovanni Spitale, Federico Germani, Ana Marušić (Croatia, Switzerland)

Consistency and Completeness of Retractions in Public Health Research on COVID-19

Caitlin Bakker, Erin Reardon, Sarah Brown, Nicole Theis-Mahon, Sara Schroter, Lex Bouter, Maurice Zeegers (Canada, Netherlands, United Kingdom, United States)

Paper Mills

Analysis of Cancer Research Discussion Text and References in High-Impact Factor Journals for Possible Indicators of Paper Mills

Annie Whamond, Adrian Barnett, Jennifer Byrne (Australia)

Screening Articles for Tortured Phrases With a Regular Expressions-Based Detector

Alexandre Clausse, Guillaume Cabanac, Pascal Cuxac, Cyril Labbé (France)

Peer Review

Librarian and Information Specialist Perceptions of Peer Reviewing Systematic Reviews

Melissa Rethlefsen, Carrie Price, Sara Schroter (United Kingdom, United States)

An Agent-Based Modeling Approach for Evaluating Interventions to Optimize Peer Review

Francesco De Pretis, Abdelghani Maddi, Ahmad Yaman Abdin (France, Germany, United States)

Evidence of Use of Template-Based Peer-Review Reports and Concern About Review Mills

Cyril Labbé, Gilles Hubert, Wendeline Swart, Guillaume Cabanac (France)

Virtual

Research Culture Influences in Peer Review: A Targeted Thematic Analysis of Current Challenges in Peer Review

Lesley Uttley, Yuliang Weng, Louise Falzon (United Kingdom)

Motivations to Participate in the Peer Review Process at the *Journal of Urology*

Anne Dudley, George Koch, Kyle Rose, Roei Golan, Jennifer Regala, Casey Seideman, Amanda North, Kevin Koo, Kevan Sternberg, Gina Badalato, Benjamin Dropkin, Nicholas Chakiryan, Robert Siemens, Peter Clark, Andrew Harris (Canada, United States)

Review of Proposals Submitted to Elsevier's Peer Review Workbench

Bahar Mehmani, Silvia Dobre, Ramadurai Petchiappan (Netherlands)

Peer Review Process and Models

Efficiency of Author Anonymization in Peer Review

Markus Heinemann, Andreas Boening, Kazunori Okabe, Jessica Bogensberger, Zulfugar Taghiyev (Germany, Japan)

Manuscript Submissions Following Implementation of Guaranteed Peer Review

Yurong Fei Bloom, Matthew Welch (United States)

Editor Initial Manuscript Review: A Masked Pilot Study

Douglas Novins, Mary Billingsley, Robert Althoff (United States)

Optimizing Proposal Assignments in the Distributed Peer Review System of the World's Largest Radio Telescope Observatory

Andrea Corvillon, John Carpenter, Nihar Shah (Chile, United States)

AI-Augmented Peer Review, Collaboration Dynamics, and Human Reviewer Performance

Ashia Livaudais, Dmitri Iourovitski (United States)

Virtual

Use of a 3-Round Modified Delphi Format to Support Robust, Rapid Peer Review

Sean Hays Tyler Carneal, Christopher Kirman (United States)

Predatory Journals

Persistence and Indexing of Predatory Journals and Publishers: A Follow-Up Evaluation of Beall's List

Pravin Bolshete, Madhulika Bolshete, Priyanka Mate (India)

Virtual

Inclusion of Randomized Controlled Trials Published by Potentially Predatory Journals in Anesthesiology Systematic Reviews: A Cross-Sectional Study
Julián Velásquez Paz, Andrés Zorrilla Vaca, Markus Klimek, Jose Calvache (Colombia, Netherlands, United States)

Preprints

Preprint Policies in Ecology and Evolutionary Biology Journals
Marija Purgar, Edward Ivimey-Cook, Antica Culina, Joshua Wallach (Croatia, United Kingdom, United States)

Authors' Journeys From Preprints With *The Lancet* on SSRN to Publication
Sherrie Kelly, Clare Stone, Catherine Fiscus, Ashlie Jackman-Juler, Miriam Lewis Sabin (Canada, United Kingdom, United States)

Preregistration of Studies

Impact of ICMJE Trial Registration Policy at 20 Years
Julianne Nelson, Tony Tse, Swapna Mohan, Yvonne Puplampu-Dove (United States)

Effectiveness of Preregistration in Psychology
Olmo Van den Akker, Marcel Van Assen, Marjan Bakker, Jelte Wicherts (Netherlands)

Registration of Observational Studies of Interventions: Prevalence, Characteristics, and Journal Policies
Cecilie Jespersen, Zexing Song, An-Wen Chan, Asbjørn Hróbjartsson (Canada, Denmark)

International Registered Reports Identifiers (IRRIDs): 7 Years of Experiences
Gunther Eysenbach (Canada)

Virtual

Outcome Switching in Observational Studies of Interventions: Comparison of Registration Records and Published Articles
Zexing Song, Cecilie Jespersen, Asbjørn Hróbjartsson, S. Joseph Kim, Rob Fowler, Peter Austin, An-Wen Chan (Canada, Denmark)

Quality of Reporting

Assessing the Quality and Timeliness of Results Reporting for Clinical Trials on Antimicrobial Agents
Megan Curtin, Allisun Wiltshire, Brix Kowalski, Maximilian Siebert (United States)

Reporting Study Design in Korean Medical Journal Articles
Soo Young Kim, Sue Kim, Hyun Jung Yi (South Korea)

Comparative Analysis of Expert, Clinician, and Consumer Interactions With Summary of Findings Tables: A Quasi-Experimental Study
Nina Vitlov, Nensi Bralić, Tina Poklepović-Peričić, Daniel Garcia-Costa, Emilia Lopez-Iñesta, Elena Álvarez-García, Francisco Grimaldo, Ana Marušić (Croatia, Spain)

Replication and Impact of Positive Secondary Findings in Negative or Neutral Cardiovascular Trials
Sina Rashedi, Farbod Zahedi Tajrishi, Ashkan Hashemi, Isaac Dreyfus, Nicholas Varunok, John Burton, Seng Chan You, Bjorn Redfors, Gregory Piazza, Joshua Wallach, Lesley Curtis, Sanjay Kaul, David Cohen, Roxana Mehran, Mitchell Elkind, Flavia Geraldes, Joseph Ross, Jane Leopold, Harlan Krumholz, Gregg Stone, Behnood Bikdeli (South Korea, United Kingdom, United States)

Public Availability of Randomized Clinical Trial Protocols: A Repeated Cross-Sectional Study
Christof Schönenberger, Malena Chiaborelli, Ala Heravi, Lukas Kübler, Pooja Gandhi, Zsuzsanna Kontar, Julia Hüllstrung, Mona Elalfy, Jan Glasstetter, Dmitry Gryaznov, Belinda von Niederhäusern, Anette Blümle, Jason Busse, Szimonetta Lohner, Sally Hopewell -Reviewer, Matthias Briel, Benjamin Speich (Canada, Germany, Hungary, Switzerland, United Kingdom)

Adherence to the WHO Statement on Public Disclosure of Clinical Trial Results by Trials Published in High Impact Factor Journals
Carolina Grana, Lina Ghosn, Carolina Riveros, Philippe Ravaud, Isabelle Boutron (France)

Prevalence of Prospective Registration and Primary Outcome Discrepancies in Recently Published Randomized Controlled Trials
Ioana Alina Cristea, Florian Naudet, Guillaume Cabanac, John Ioannidis (France, Italy, United States)

Virtual

Developing a Harmonized Approach for Reporting Irradiation Protocols and Methods for Research Using X-ray Irradiators

Warren Stern, Ioanna Iliopulos, Christopher Boyd, Sidra Zia (United States)

Reporting Guidelines

Challenges in Achieving Uptake and Journal Endorsement of the ACcurate COnsensus Reporting Document (ACCORD) Guideline

Christopher Winchester, Mark Rolfe, William Gattrell, Patricia Logullo, Kieth Goldman, Amy Price, Paul Blazey, Esther van Zuuren, Niall Harrison (Canada, Netherlands, United Kingdom, United States)

Transparent Reporting of Observational Studies Emulating a Target Trial: The TARGET Guideline

Aidan Cashin, Harrison Hansford, Miguel Hernán, Sonja Swanson, Hopin Lee, Matthew Jones, Issa Dahabreh, Barbra Dickerman, Matthias Egger, Xabier Garcia-Albeniz, Robert Golub, Nazrul Islam, Sara Lodi, Margarita Moreno-Betancur, Sallie-Anne Pearson, Sebastian Schneeweiss, Melissa Sharp, Jonathan Sterne, Elizabeth Stuart, James McAuley (Australia, Ireland, Spain, Switzerland, United Kingdom, United States)

Author Practices and Experiences With PRISMA-P 2015

Mette B. Engmose, An-Wen Chan, Kerry Dwan, Carsten Hinrichsen, Asbjørn Hróbjartsson, David Moher, Larissa Shamseer, Lesley Stewart, Camilla Nejstgaard (Canada, Denmark, United Kingdom)

Development of a Reporting Guideline on Health Equity in Observational Research (STROBE-Equity)

Vivian Welch, Catherine Chamberlain, Peter Craig, Luis Gabriel Cuervo, Omar Dewidar, Holly Ellingwood, Elizabeth Ghogomu, Billie Jo Hardy, Tanya Horsley, Sonya Faber, Condy Feng, Damian Francis, Sarah Funnell, Alison Krentel, Janet Jull, Elizabeth Kristjansson, Julian Little, Lovelin Lum Niba, Tamara Kredo, Zack Marshall, Lawrence Mbuagbaw, Michael Mahande, Stuart Nicholls, Miriam Nkangu Nguilefem, Ekwaro Obuku, Oyekola Oloyede, Ebenezer Owusu-Addo, Kevin Pottie, Jacqueline Ramke, Alison Riddle, Anita Rizvi, Janet Hatcher Roberts, Larissa Shamseer, Melissa Sharp, Janice Tufte, Peter Tugwell, Xiaoqin Wang, Laura Weeks, Charles Wisonge, Luke Wolfenden, Taryn Young (Australia, Cameroon, Canada, Ghana, Ireland, Kenya, New Zealand, South Africa, Uganda, United Kingdom, United States)

Reproducibility

Prevalence of Reproducible Health Sciences Research: A Systematic Review and Meta-Analysis

Niklas Bobrovitz, Harriet Ware, Corson Johnstone, Juliane Kennett, Stephana Moss, Liam Whalen-Browne, Faizan Khan, Benjamin Fletcher, Daniel Niven, Henry Stelfox (Canada)

Factors Associated With the Reproducibility of Health Sciences Research: A Systematic Review and Evidence Gap Map

Stephana Julia Moss, Juliane Kennett, Jeanna Parsons Leigh, Niklas Bobrovitz, Henry T. Stelfox (Canada)

Perception of Open Science Practices on Reproducibility Among Reviewers of Grant Proposals at Research Funding Organizations

Ayu Putu Madri Dewi, Nicholas J. DeVito, Gowri Gopalakrishna, Inge Stegeman, Mariska Leeflang (Netherlands, United Kingdom)

Virtual

Testing Computational Reproducibility Review in Editorial Workflows of Academic Journals: A Randomized Controlled Trial From the European iRISE project

Laura Caquelin, Rachel Heyard, Stephanie Zellers, Hanno Würbel, Gustav Nilsonne (Finland, Sweden, Switzerland)

Research Methods

Characterizing Adverse Event Methods Reported in ClinicalTrials.gov and Publications

Kyungwan Hong, Mark Basista, Tony Tse (United States)

Bias in Machine Learning Associated With Weak Baselines, Data Leakage, and Inadequate Measures Reporting

Randall Ellis, Chirag J. Patel (United States)

Methodological Guidance for Individual Participants Data Meta-Analyses: A Systematic Review

Edith Otalike, Michael Clarke, Ngianga-Bakwin Kandala, Joel Gagnier (Canada, Northern Ireland)

Reporting of Confounder Selection in Observational Studies in High Impact Medical and Epidemiological Journals, 2003-2023

Luis Correia, Rafael Mascarenhas, Felipe de Menezes, Jeronimo Oliveira Júnior, Marcus Almeida, Caio Azevedo, Naieli de Andrade, Viola Vaccarino, Joseph Ross, Joshua Wallach (Brazil, United States)

Non-Inferiority vs Superiority Trials in Cardiovascular Research: Trends, Success in Meeting the Primary Outcome, and Online Engagement

Ashkan Hashemi, Isaac Dreyfus, Nicholas Varunok, John Burton, Sina Rashedi, Farbod Zahedi Tajrishi, Behnood Bikdeli (United States)

Virtual

Estimation of an Upper Limit on the Maximum Effect That Can Be Detected in Randomized Trials of Cancer Therapeutics

Benjamin Djulbegovic, Iztok Hozo, Renata Iskander, Austin Parish, Jonathan Kimmelman, John Ioannidis (Canada, United States)

Performance and Practicality of a Randomized Clinical Trial Classifier in Systematic Literature Reviews vs a Traditional Approach

Ambar Khan, Ania Bobrowska, Chloe Coelho, Hannah Frost, Swati Kumar, Hannah Russell, Anna Noel-Storr, Molly Murton (United Kingdom)

Retractions

An Audit and Feedback Intervention to Reduce Inappropriate Citation of Retracted Literature in the Pain and Anesthesiology Fields

Michael Ferraro, Aidan Cashin, Amanda Williams, Emma Fisher, Gavin Stewart, Christopher Eccleston, Neil O'Connell (Australia, United Kingdom)

Prevalence of and Reasons for Retractions of Traditional Chinese Medicine Research Publications by Authors From Mainland China in International Peer-Reviewed Journals

Jing Cui, Nan Yang, Kexin Ji, Dingran Yin, Chen Shen, Zhaolan Liu, Han Tan, Yaxin Sun, Zhaoqi Huo, Shuo Liu, Huiyu Wang, Xintong Zhang, Jing Guo, Yufei Wang, Xiaoqi Ren, Vincent Chung, Jianping Liu (China, Hong Kong)

Citation Context Analysis of Retracted Articles: Leveraging Retraction Reasons to Track Unreliability in Citing Literature

Yagmur Ozturk, Frédérique Bordignon, Cyril Labbé, François Portet (France)

Evaluating Approaches for Identifying Retracted Articles and Retraction Notices in Systematic Review Searching

Caitlin Bakker, Erin Reardon, Nicole Theis-Mahon, Sara Schroter, Lex Bouter, Maurice Zeegers (Canada, Netherlands, United Kingdom, United States)

Postretraction References in Biomedical and Clinical Sciences

Kathryn Weber-Boer, Guillaume Cabanac (France, United Kingdom)

Retracted Publications Referenced in Clinical Guidelines

Kathryn Weber-Boer, Guillaume Cabanac, Lonni Besançon (France, Sweden, United Kingdom)

Retraction of Systematic Reviews and Clinical Practice Guidelines

Ivan D. Florez, Alberto Henriquez, Andres Estupinan-Bohorquez (Colombia)

Social Media

Video and Social Media Performance at a Surgical Journal with Video Journal Clubs

Caden Seraphine, Abigail Chambers, Susan Galandiuk (United States)

Virtual

Altmetric Footprint of Retracted and Corrected Publications: The Role of Misinformation and Disinformation

Ashraf Maleki, Niina Sormanen, Kim Holmberg (Finland)

Plenary Session Invited Talks

Drummond Rennie Lecture

Forward to the Past—Making Contributors Accountable

Ana Marušić[1]

Importance Authorship is the currency of academia and research, but it is not only about recognition. It brings accountability and transparency and can have ethical and legal implications.

Observations Starting from the call for contributors to replace authors by Drummond Rennie and his colleagues almost 30 years ago,[1] the evolution of scientific authorship and contributorship models will be presented. Current challenges to authorship will also be explored, from paper mills to artificial intelligence.

Conclusions Authorship has not been replaced by contributorship and will need a systemic approach to balance scientific credit, career advancement, and ethical responsibility.

Reference

1. Rennie D, Yank V, Emanuel L. When authorship fails: a proposal to make contributors accountable. *JAMA*. 1997;278(7):579-585. doi:10.1001/jama.278.7.579

[1]University of Split School of Medicine, Split, Croatia; ana.marusic@mefst.hr.

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.

The Douglas G. Altman Lecture

Does the Journal Article Have a Future?

Malcolm Macleod[1]

Importance From the perspective of the researcher, there seems to be too much research, of often questionable quality, in too many journals, published following too little scrutiny.

Observations It has become more difficult to distinguish between intrusive email solicitations from "proper" publishers and those from predatory publishers. While many employed in publishing have the most noble of intentions, the profits of publishing houses—profit drawn largely from our research budgets—continue to soar. At the same time, alternative routes of dissemination, such as preprint servers, now allow research findings to be made public, if not published, at minimal cost. Loud protestations from journals that preprints are of inferior quality because they have not been peer reviewed would carry more weight if the process of peer review was a reliable guarantee of quality and the provenance of the data presented. Despite several decades of opportunity to increase the value of journal articles by investing in these processes, most publishers have not done so. Many in my own community have been complicit in the generation of "research for publication" rather than "research for the advancement of knowledge" (where publication is a part of the process, but not the primary objective). Drivers for these behaviors come from across the research ecosystem, and the focus on journal publication as the cardinal researcher output for evaluation serves as the major enabler of exploitative publisher behaviors.

Conclusions If the journal article is to have a future, this extractive behavior must stop. The added value of peer review needs to be clear and apparent, and the quality of published work must be such that research users can trust the provenance of what is reported; with sufficient information, data, and code provided so that they can check that provenance if they wish. The quality of published work is improving, in some respects, at some journals, and those who have driven these changes, often in the face of internal opposition, should be celebrated. I will close by outlining some further avenues for improvement.

[1]Edinburgh Neuroscience, University of Edinburgh, Edinburgh, Scotland, malcolm.macleod@ed.ac.uk.

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

Journal Prestige Can and Should Be Earned

Simine Vazire[1]

Importance Scientific institutions, including journals, should work to earn trust from the scientific community and from the public. With prominent threats to trust in science around the world, it is especially important to make clear why scientific institutions are worthy of trust. Too often, journals' reputations are unearned—based on flawed metrics, such as impact factors—or simply the inertia of prestige. But journal prestige can and should be earned.

Observations There are many things journals can do to give the community concrete, verifiable indicators of their priorities. First, journals can thoroughly vet the articles they

publish to ensure they are accurate, well-calibrated, and transparently reported. Second, journals can invest in postpublication critique and correction so that when they inevitably publish some things that are wrong or miscalibrated, they correct themselves. Third, journals can encourage audits of their published articles, such as those conducted by the Institute for Replication.

Conclusions Scientific journals will sometimes publish research that turns out to be wrong. That is not a reason to distrust journals or the peer review process. But journals do have an obligation to take reasonable steps to vet research before it is published and to make those efforts visible and verifiable to readers. Journals that invest in and facilitate both prepublication and postpublication quality checks, error detection, and correction are the ones that deserve the most trust and the most prestige.

[1]Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Australia, simine@gmail.com.

Conflict of Interest Disclosures None reported.

A Singular Disruption of Scientific Publishing—AI Proliferation and Blurred Responsibilities of Authors, Reviewers, and Editors

Isaac Kohane[1,2]

Importance The full cycle from manuscript creation through review and then distribution is currently a heavily human-intensive process. If the continually increasing capabilities of artificial intelligence (AI) can replace much of that human effort, the economics, quality, culture, and reproducibility of scientific publishing will be dramatically affected internationally.

Observations AI is entering into the reviewing process in a few journals. This is going to accelerate rapidly. It will perform higher-quality reviews than a larger majority of the current batch of reviewers. Authorship, including what used to be considered as intellectual contributions, will be broadly extended to AI. This may have as much impact on the consequences for institutional promotion as for the publication process. Reviews for reproducibility and accuracy are highly likely to become ubiquitous comprehensive processes rather than the artisanal passion project of a few. Gamesmanship of citation networks as currently practiced will become easily unraveled.

Conclusions AI is going to affect all aspects of the scientific process. Because it impinges on roles for which we attributed respect, financial reward, and trust, the current publishing process is going to be transformed within the decade. Who will lead the transformation depends on a few critical decisions we make in the next 1 to 2 years.

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, US, isaac_kohane@hms.harvard.edu; [2]*NEJM AI*, Waltham, MA, US.

Conflict of Interest Disclosures None reported.

Plenary Session Abstracts

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

All Abstracts are available to post comments and questions during the meeting via the meeting platform at underline.io/events/476/reception

Wednesday, September 3

Author and Reviewer Use of AI

Authors Self-Disclosed Use of Artificial Intelligence in Research Submissions to 49 BMJ Group Biomedical Journals

Isamme AlFayyad,[1] Maurice Zeegers,[2] Lex Bouter,[3] Helen Macdonald,[4] Sara Schroter[4]

Objective Recent surveys report that a high percentage (50%-76%) of researchers use artificial intelligence (AI) in their research.[1,2] In April 2024, BMJ Group mandated submitting authors to disclose and describe their use of AI.[3] We analyzed the frequency of self-disclosed use of AI in research manuscripts submitted to 49 BMJ Group biomedical journals and identified types of AI tools used and the tasks they assisted with. We also compared the characteristics of manuscripts disclosing AI use compared with those not disclosing it.

Design A cross-sectional analysis of original research submitted between April and November 2024 was conducted and reported using the STROBE reporting guideline. Main study outcomes were frequency of self-disclosed use of AI, types of AI tools used, and tasks they assisted with. Characteristics of submissions (region of submitting author, number of authors, acceptance rate, impact factor, general medical/specialty journal, peer review model) were extracted from journals' manuscript tracking systems and websites. χ^2 test was used to compare proportions of these characteristics between submissions disclosing and not disclosing AI use. Factors associated with disclosing AI use were assessed using binomial logistic regression. A *P* value threshold for significance was set at .05.

Results There were 25,114 eligible submissions from Asia (13,505 [53.8%]), Europe (6523 [26.0%]), North America (2795 [11.1%]), Africa (1196 [4.8%]), Oceania (708 [2.8%]), and South America (387 [1.5%]). A total of 1431 of 25,114 submissions (5.7%) disclosed AI use. The most common types of AI tools used were generative AI chatbots (812/1431 [56.7%]), writing assistant tools (182 [12.7%]), and visual/image processing tools (38 [2.7%]) (**Table 25-1013**). The majority of submitting authors who disclosed AI use reported using it to improve the quality of their writing (1248/1431 [87%]), and 28% (n = 399) reported using it for other purposes, including translation (107 [7.5%]), analyzing or collecting data (44 [3.1%]), generating data and output (87 [6.1%]), code writing (15 [1.0%]), image processing (36 [2.5%]), literature searches (49 [3.4%]), and managing references (8 [0.6%]). The percentage of submissions with disclosed AI use varied significantly by region (highest in Europe: 31.0%; lowest in Oceania: 1.0%; $P < .001$). Submissions disclosing AI use had a lower mean number of authors (8.39 vs 8.90; $P = .005$). Authors from Europe (odds ratio [OR], 3.61 [95% CI, 2.11-6.18]), North America (OR, 2.30 [95% CI, 1.32-4.02]), South America (OR, 4.93 [95% CI, 2.62-9.28]), Asia (OR, 2.87 [95% CI, 1.68-4.89]), and Africa (OR, 3.34 [95% CI, 1.92-6.08]) were significantly more likely to disclose AI use than those from Oceania ($P < .05$ for all). Conversely, each additional author reduced disclosure odds by 1% (OR, 0.99 [95% CI, 0.97-0.99]). Other characteristics of the journal (acceptance rate, impact factor, general medical/specialty journal, and peer review model) were not associated with AI use disclosure.

Conclusions The percentage of submitted articles with self-disclosed AI use was significantly lower in this study than what has been reported in recent surveys of researchers about their general use. Submitting authors may be underreporting their use of AI.

Table 25-1013. Frequency of Use of Artificial Intelligence (AI) Types (N = 1431)

Type	No. (%)[a]
AI chatbots	812 (56.7)
Writing assistant tools	182 (12.7)
Visual/image processing tools	38 (2.7)
Evidence synthesis tools	28 (2)
Predictive analytics models	21 (1.5)
AI-powered translators	19 (1.3)
AI-powered data analysis tool	9 (0.6)
AI-powered markdown editor	7 (0.5)
Other	7 (0.5)
Automatic speech recognition systems	4 (0.3)
Type not disclosed	449 (31.4)

[a]The sum of percentages exceeds 100% because authors may have reported use of more than 1 AI tool or more than 1 use.

References

1. How are researchers responding to AI? Oxford University Press. May 23, 2024. Accessed September 1, 2024. https://corp.oup.com/news/how-are-researchers-responding-to-ai/

2. Insights 2024: Attitudes Toward AI. Elsevier. 2024. Accessed September 1, 2024. https://www.elsevier.com/en-gb/insights/attitudes-toward-ai

3. Macdonald H, Abbasi K. Riding the whirlwind: *BMJ*'s policy on artificial intelligence in scientific publishing. *BMJ*. 2023;382:1923. doi:10.1136/bmj.p1923

[1]Research Center, King Fahad Medical City, Saudi Arabia, ialfayyad@gmail.com; [2]Department of Epidemiology, Care and Public Health Research Institute, Maastricht University, the Netherlands; [3]Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Department of Philosophy, Faculty of Humanities, Vrije Universiteit Amsterdam, the Netherlands; [4]*BMJ,* London, UK.

Conflict of Interest Disclosures This research is part of an ongoing PhD collaboration between The BMJ and the team at Meta-Research at Maastricht University (UM) on the responsible conduct of publishing scientific research. *The BMJ* is published by BMJ Group, a wholly owned subsidiary of the British Medical Association. UM is a public legal entity in the Netherlands. This study is part of Isamme AlFayyad's self-funded BMJ/UM PhD. No exchange of funds has taken place for this research project. All authors express their own opinions and not necessarily that of their employers. Helen Macdonald and Sara Schroter are full-time employees at BMJ Publishing Group. No other disclosures were reported.

Additional Information The protocol of this research study was registered in Open Science Framework (osf.io/uvqms).

Artificial Intelligence Use and Acknowledgment in Medical Research Writing Among Chinese Scholars

Yang Zhang,[1,2] Mengyuan Duan,[1,2] Guoguang Zhao,[3] Penghu Wei,[2,3] Xiuyuan Hao[4]

Objective Given the increasing role of artificial intelligence (AI) in academic writing, it is important to understand how these tools are integrated into research practices. This study aimed to explore the use of AI in academic writing by Chinese medical scholars and to identify potential factors influencing their adoption and perceptions of AI.

Design A total of 378 peer reviewers of *Chinese Medical Journal* were invited to participate in a survey using an electronic questionnaire via email between January 2, 2025, and January 12, 2025. The questionnaire included questions about AI tool use in research paper writing, reasons for nonuse, stages of use, experience of declaration, and perception of effects of AI use. The survey also explored what role AI should have and where researchers believe AI should be acknowledged in publications. The Fisher exact test was used to identify significant differences between subgroups based on age and use.

Results Of the 378 reviewers invited, 159 (42.1%) provided valid responses. Among the respondents, 93 (58.5%) reported using AI tools in their research paper writing, and 66 (41.5%) did not (**Table 25-0878**). There was a significant difference in declared use of AI across age groups, with younger researchers more likely to use AI tools (76.1% of those ≤30 years, 59.1% of those 31-45 years, 27.8% of those 46-60 years, and 14.3% of those >60 years; $P < .001$). The primary reasons for not using AI tools included copyright or ethical issues (53.0%), concerns about inaccuracy (47.0%), and unfamiliarity with AI tools (40.9%). Notably, there were still 27.3% non-AI users considering using AI when technology matures. The most common stages at which AI tools were used was after writing for language translation and polishing (90.3%), followed by idea generation before writing (16.1%) and structural organization during writing (15.1%). However,

Table 25-0878. Questionnaire Results on the Use and Acknowledgment of AI Tools in Research Manuscript Writing Among Chinese Medical Scholars (N = 159)

Question and response	No. (%)
Q1: Age of participants, y	
A. ≤30	46 (28.9)
B. 31-45	88 (55.3)
C. 46-60	18 (11.3)
D. >60	7 (4.4)
Q2: Do you use AI tools when writing research papers?	
A. Yes	93 (58.5)
B. No	66 (41.5)
Q3: Why don't you use AI tools for research paper writing? (only for No response to Q2)[a]	
A. Unfamiliar with or unable to use AI tools	27 (40.9)
B. Inaccurate or unsuitable generated content	31 (47.0)
C. Concerns about copyright or ethical issues	35 (53.0)
D. Consider using it once AI technology matures	18 (27.3)
E. Unable to find a suitable AI tool	20 (30.3)
F. Other	2 (3.0)
Q4: At which stages do you primarily use AI tools? (only for Yes response to Q2)[a]	
A. Before writing: idea generation and planning	15 (16.1)
B. During writing: structure organization and data analysis	14 (15.1)
C. After writing: language translation, and polishing	84 (90.3)
D. Other	3 (3.2)
Q5: Any relevant declarations of using AI in past submissions? (only for Yes response to Q2)	
A. Yes	27 (29.0)
B. No	66 (71.0)
Q6: Do you think AI tools are helpful for your research paper? (only for Yes response to Q2)	
A. Very helpful	56 (60.2)
B. Somewhat helpful	37 (39.8)
C. Negative impact	0
Q7: What role or where should AI tools be acknowledged in the paper?	
A. As a co-author of the paper	6 (3.8)
B. In the methods section	20 (12.6)
C. In the conflict of interest section	8 (5.0)
D. In a separate AI writing declaration section	98 (61.6)
E. Should not be acknowledged	27 (17.0)

Abbreviation: AI, artificial intelligence.

[a]Participants could select multiple responses.

only 29.0% of participants declared AI use in past submissions. A total of 60.2% found the tools to be very helpful, while none considered them to have negative impact. The preferred roles and acknowledgement of AI tools in papers included a separate AI writing declaration section (61.6%), not acknowledging AI at all (17.0%), reporting in the methods section (12.6%), reporting in the conflict of interest section (5.0%), and including AI as a co-author (3.8%), which did not significantly differ between AI users and nonusers; nor did these views differ significantly by age.

Conclusions In this study, nearly 60% of Chinese medical scholars reported use of AI tools in writing and preparing papers, particularly for language translation and polishing. However, many reported concerns about the accuracy, ethics, and maturity of AI technology. Current AI guidelines in academic publishing lack a unified standard, particularly in usage scope and enforcement. Some journals lack mandatory disclosure policies for AI use, standardized templates for reporting, or requirements to submit archival copies of original data or content before AI use, which might account for incomplete reporting or failure to report use of AI. Further research is needed to assess the long-term impact of AI on academic writing.

[1]National Center for Neurological Disorders, Xuanwu Hospital Capital Medical University, Beijing, China; [2]Editorial Office, *Brain Network Disorders*, Xuanwu Hospital Capital Medical University, Beijing, China; [3]Department of Neurosurgery, Xuanwu Hospital Capital Medical University, Beijing, China; [4]Editorial Office, *Chinese Medical Journal*, Chinese Medical Association Publishing House, Beijing, China, haoxiuyuan@cma.org.cn.

Conflict of Interest Disclosures None reported.

AdditionaI Information ChatGPT (version GPT-4o mini; OpenAI) was used on January 20, 2025, for language translation, grammar checking, and manuscript polishing to ensure the clarity and coherence of this abstract.

Factors Associated With Author and Reviewer Declared Use of AI in Medical Journals

Roy H. Perlis,[1,2] Annette Flanagin,[3] Jacob Kendall-Taylor,[3] Michael Berkwits,[4] Kirsten Bibbins-Domingo[3]

Objective The scientific community is divided about the appropriate role of artificial intelligence (AI) in manuscript preparation.[1] In 2023, JAMA Network journals began requiring authors and reviewers to answer questions about use of AI to create or assist in the creation or editing of submitted manuscripts or with preparation of reviews.[2] This study was conducted to assess author and reviewer declarations of such use.

Design Study of manuscripts and peer reviews submitted to 13 JAMA Network journals from September 2023 to May 2025. Numbers and proportions of manuscripts with corresponding authors and peer reviewers declaring use of AI were examined, along with factors associated with use (eg, journal, country, article type). For authors' use, reasons, type of AI, and editorial decisions were also assessed; for reviewers' use, turnaround time and ratings of reviews were assessed. Where sample sizes were sufficient, logistic regression models were run, with use of AI as the dependent variable and article type, journal, and review outcome as independent variables. A 2-sided uncorrected *P* value of <.05 was considered the threshold for statistical significance. This study was deemed exempt from ethical approval by the Mass General Brigham Institutional Review Board.

Results Of 82,829 manuscripts submitted during the study period, 2257 (2.7%) declared AI use. Authors' use of AI increased from 1.6% in September 2023 to 4.2% in May 2025 (r^2 = .34; P < .01). **Table 25-1115** contrasts manuscripts with or without disclosed AI use by author country, journal, manuscript type, study type, editorial decision, AI use categories, and common AI models used. Author AI use was highest in *JAMA Network Open* (4.1%) and *JAMA* (3.1%) and lowest in *JAMA Dermatology* (1.6%) and was highest among authors from Taiwan (7.2%). Letters to the Editor (odds ratio [OR] 1.63, 95% CI 1.37-1.94) and Viewpoints (OR 1.44, 95% CI 1.21-1.72) were more likely to involve AI use than Original Investigations. The most common reported reasons for AI use by authors were for language and editing (50.0%), data analysis and statistics (11.8%), and content generation and drafting (8.2%). ChatGPT was the most common AI model used by authors (62.9%). Mean (SD) reviewer turnaround times were 11.8 (8.5) days for those who did not disclose AI use and 12.6 (7.6) days for reviewers who disclosed AI use. Reviewer use of AI was highest for *JAMA Network Open* (0.9%) and *JAMA Psychiatry* (0.8%) and among reviewers from the UK (2.7%).

Conclusions Among manuscripts submitted to 13 major medical journals, author declaration of AI use to create or edit content was low but more than doubled during the study period and varied by journal and article type. Very few reviewers disclosed use of AI, possibly because of guidance forbidding such use.[3] Limitations include small numbers of reported AI use, which limited comparisons for some factors, and the likelihood of underreporting of AI use.

References

1. Kwon D. Is it OK for AI to write science papers? *Nature* survey shows researchers are split. *Nature*. 2025;641:574-578. doi:10.1038/d41586-025-01463-8

2. Flanagin A, Pirracchio R, Khera R, Berkwits M, Hswen Y, Bibbins-Domingo K. Reporting use of AI in research and scholarly publication—JAMA Network guidance. *JAMA*. 2024;331(13):1096-1098. doi:10.1001/jama.2024.3471

3. Flanagin A, Kendall-Taylor J, Bibbins-Domingo K. Guidance for authors, peer reviewers, and editors on use of AI, language models, and chatbots. *JAMA*. 2023;330(8):702-703. doi:10.1001/jama.2023.12500

[1]JAMA+ AI and *JAMA Network Open,* Chicago, IL, rperlis@mgh.harvard.edu; [2]Department of Psychiatry, Mass General Brigham, Boston, MA, US; [3]*JAMA* and the JAMA Network, Chicago, IL, US; [4]Centers for Disease Control and Prevention, Atlanta, GA, US.

Conflict of Interest Disclosures Roy H. Perlis receives payment for service as a scientific advisor to Circular Genomics, Genomind,

Table 25-1115. Author and Reviewer Disclosure of Artificial Intelligence (AI) Use in 13 Medical Journals, September 2023-May 2025

	Authors, No. (%)			Reviewers, No. (%)		
	No AI use (n = 80,572)	AI use (n = 2257)	Total (N = 82,829)	No AI use (n = 42,361)	AI use (n = 250)	Total (N = 42,611)
Journal						
JAMA Cardiology	4107 (98.2)	74 (1.8)	4181 (5.0)	1702 (99.6)	6 (0.4)	1708 (4.0)
JAMA Dermatology	5531 (98.4)	91 (1.6)	5622 (6.8)	2149 (99.5)	10 (0.5)	2159 (5.1)
JAMA Internal Medicine	6217 (97.7)	144 (2.3)	6361 (7.7)	1954 (99.5)	10 (0.5)	1964 (4.6)
JAMA	1,5091 (96.9)	481 (3.1)	15,572 (18.8)	8253 (99.6)	30 (0.4)	8283 (19.4)
JAMA Health Forum	2865 (97.8)	65 (2.2)	2930 (3.5)	1478 (99.7)	4 (0.3)	1482 (3.5)
JAMA Network Open	15,904 (95.9)	680 (4.1)	16,584 (20.0)	13,651 (99.1)	125 (0.9)	13,776 (32.3)
JAMA Neurology	4668 (97.9)	101 (2.1)	4769 (5.8)	2171 (99.5)	12 (0.5)	2183 (5.1)
JAMA Oncology	5545 (97.6)	135 (2.4)	5680 (6.9)	1780 (99.4)	10 (0.6)	1790 (4.2)
JAMA Ophthalmology	4137 (97.8)	94 (2.2)	4231 (5.1)	1110 (99.5)	6 (0.5)	1116 (2.6)
JAMA Otolaryngology-Head & Neck Surgery	2639 (98.0)	54 (2.0)	2693 (3.3)	1166 (99.6)	5 (0.4)	1171 (2.7)
JAMA Pediatrics	5198 (97.9)	114 (2.1)	5312 (6.4)	1760 (99.5)	9 (0.5)	1769 (4.2)
JAMA Psychology	3801 (97.2)	110 (2.8)	3911 (4.7)	1920 (99.2)	16 (0.8)	1936 (4.5)
JAMA Surgery	4869 (97.7)	114 (2.3)	4983 (6.0)	3267 (99.8)	7 (0.2)	3274 (7.7)
Country[a]						
US	34,592 (97.8)	770 (2.2)	35,362 (42.7)	32,064 (99.7)	99 (0.3)	32,163 (75.5)
China	19,324 (97.7)	455 (2.3)	19,779 (23.9)	NA	NA	NA
Canada	2750 (98.0)	57 (2.0)	2807 (3.4)	1739 (99.5)	8 (0.5)	1747 (4.1)
South Korea	2238 (97.9)	49 (2.1)	2287 (2.8)	NA	NA	NA
Taiwan	1734 (92.8)	134 (7.2)	1868 (2.3)	NA	NA	NA
UK	NA	NA	NA	1267 (97.3)	35 (2.7)	1302 (3.1)
Germany	NA	NA	NA	448 (98.7)	6 (1.3)	454 (1.1)
Australia	NA	NA	NA	637 (99.2)	5 (0.8)	642 (1.5)
Other	19,934 (96.2)	792 (3.8)	20,726 (25.0)	6206 (98.5)	97 (1.5)	6303 (14.8)
Manuscript type[a]						
Original Investigation	49,661 (97.2)	1425 (2.8)	51,086 (61.7)	30,075 (99.3)	205 (0.7)	30,280 (71.1)
Research Letter	7051 (97.3)	199 (2.7)	7250 (8.8)	5445 (99.6)	22 (0.4)	5467 (12.8)
Viewpoint	4438 (96.5)	162 (3.5)	4600 (5.6)	2251 (99.7)	7 (0.3)	2258 (5.3)
Letter to the Editor	4006 (95.7)	179 (4.3)	4185 (5.1)	NA	NA	NA
Brief Report	3165 (97.6)	78 (2.4)	3243 (3.9)	1430 (99.6)	6 (0.4)	1436 (3.4)
Narrative Review	NA	NA	NA	699 (99.7)	2 (0.3)	701 (1.6)
Other	12,251 (98.3)	214 (1.7)	12,465 (15.0)	2461 (99.7)	8 (0.3)	2469 (5.8)
Study type[a]						
Cohort	23,246 (97.6)	562 (2.4)	23,808 (28.7)	13,764 (99.3)	95 (0.7)	13,859 (32.5)
Cross-sectional	11,500 (97.2)	334 (2.8)	11,834 (14.3)	6979 (99.4)	45 (0.6)	7024 (16.5)
Randomized clinical trial	4426 (98.5)	69 (1.5)	4495 (5.4)	4946 (99.5)	27 (0.5)	4973 (11.7)
Meta-analysis	3849 (97.4)	101 (2.6)	3950 (4.8)	1741 (99.3)	12 (0.7)	1753 (4.1)
Case series	2435 (97.6)	61 (2.4)	2496 (3.0)	NA	NA	NA
Case control	NA	NA	NA	1138 (99.4)	7 (0.6)	1145 (2.7)
Other	35,116 (96.9)	1130 (3.1)	36,246 (43.8)	13,793 (99.5)	64 (0.5)	13,857 (32.5)
Manuscript decision						
Reject	66,896 (97.4)	1779 (2.6)	68,675 (82.9)	NA	NA	NA
Accept or revise	9688 (96.8)	319 (3.2)	10,007 (12.1)	NA	NA	NA
Pending/other	2955 (96.5)	107 (3.5)	3062 (3.7)	NA	NA	NA
Withdrawn	1033 (95.2)	52 (4.8)	1085 (1.3)	NA	NA	NA
Use of AI[a]						
Not specified	0	433 (19.2)	81,005 (97.8)	NA	NA	NA
Language and editing	0	1129 (50.0)	1129 (1.4)	NA	NA	NA
Data analysis and statistics	0	266 (11.8)	266 (0.3)	NA	NA	NA
Content generation and drafting	0	184 (8.2)	184 (0.2)	NA	NA	NA
Other	0	245 (10.9)	245 (0.3)	NA	NA	NA
NA	80,572 (100)	NA	NA	NA	NA	NA
Model type[a]						
ChatGPT	0	1420 (62.9)	1420 (62.9)	NA	NA	NA
Claude	0	54 (2.4)	54 (2.4)	NA	NA	NA
Gemini	0	21 (0.9)	21 (0.9)	NA	NA	NA
Bard	0	4 (0.2)	4 (0.2)	NA	NA	NA
Other or not specified	0	758 (100)	758 (33.6)	NA	NA	NA
NA	80,572 (100)	NA	NA	NA	NA	NA
Reviewer rating						
Very good or excellent	NA	NA	NA	37,675 (99.4)	224 (0.6)	37,899 (91.1)
Good or lower	NA	NA	NA	3681 (99.6)	14 (0.4)	3695 (8.9)
Reviewer turnaround time, mean (SD), d	NA	NA	NA	11.8 (8.5)	12.6 (7.6)	11.8 (8.5)

Abbreviation: NA, not applicable.

[a]Top 5 in each category in descending order by number.

and Alkermes. Annette Flanagin, Michael Berkwits, and Kirsten Bibbins-Domingo are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract.

Additional Information Llama 3.3 (Ollama) was used on June 13, 2025, to identify categories of author-declared use of artificial intelligence. Roy H. Perlis takes responsibility for the integrity of the content generated.

Quantifying and Assessing the Use of Generative AI by Authors and Reviewers in the Cancer Research Field

Daniel S. Evanko,[1] Michael Di Natale[1]

Objective This study assessed the ability to reliably detect the use of generative artificial intelligence (genAI) by authors and reviewers in the field of cancer research, quantified this usage over time, and measured the impact of a change in policy on use of genAI by reviewers.

Design In this cross-sectional study, original research manuscripts and reviewer comments submitted to 10 journals of the American Association for Cancer Research from 2021 through 2024 were analyzed for the presence of AI-generated text. The abstract, methods section, and reviewer comments were extracted from the submission system along with author and reviewer metadata. These sections were chosen because (1) the plain text was stored in the submission system, thus eliminating the need for extraction from files, and (2) we hypothesized that authors would most likely use genAI in the abstract and had observed that methods can be prone to false positives. Text was analyzed using an AI detection tool and

classified by average AI likelihood score. Discrimination between full de novo generation of text, language translation, and editing was not possible.

Results We analyzed 46,500 abstracts, 46,021 methods sections, and 29,544 reviewer comments. In 18,467 manuscripts submitted from 2021 through Q3 2022, only 7 abstracts—and no methods or reviewer comment text—were scored as highly likely to contain AI-generated text (AI likelihood score, 0.9 to 1). There were very few false positives (abstract: 0.22%, 40/18,467; methods: 0.07%, 13/18,303; reviewer comments: 0.03%, 4/12,908). Detection of AI-generated text in all 3 sources started increasing dramatically beginning in Q1 2023 after the public release of ChatGPT in November 2022 (**Figure 25-1180**). Detections in reviewer comments dropped by >50% (from 46 to 34) in Q4 2023 after the journals implemented a policy prohibiting use of genAI for peer review before resuming a linear upward trend. In aggregate, the presence of AI-generated text in 2024 varied greatly between abstracts (23%, 2749/11,959), methods text (10%, 1209/11,875), and reviewer comments (4.8%, 348/7211). There was also substantial journal-dependent variation, from 12% to 15% of abstracts at 3 journals (183/1464, 29/214, 32/214) to 28% to 33% of abstracts at 2 journals (203/716, 438/1342). Authors affiliated with institutions in English-speaking countries were less than half as likely to use genAI compared with other authors. The presence of AI-generated text was associated with a substantially higher rejection rate before peer review regardless of institution location. There was a slight association of AI-generated text with lower-rated reviewer comments.

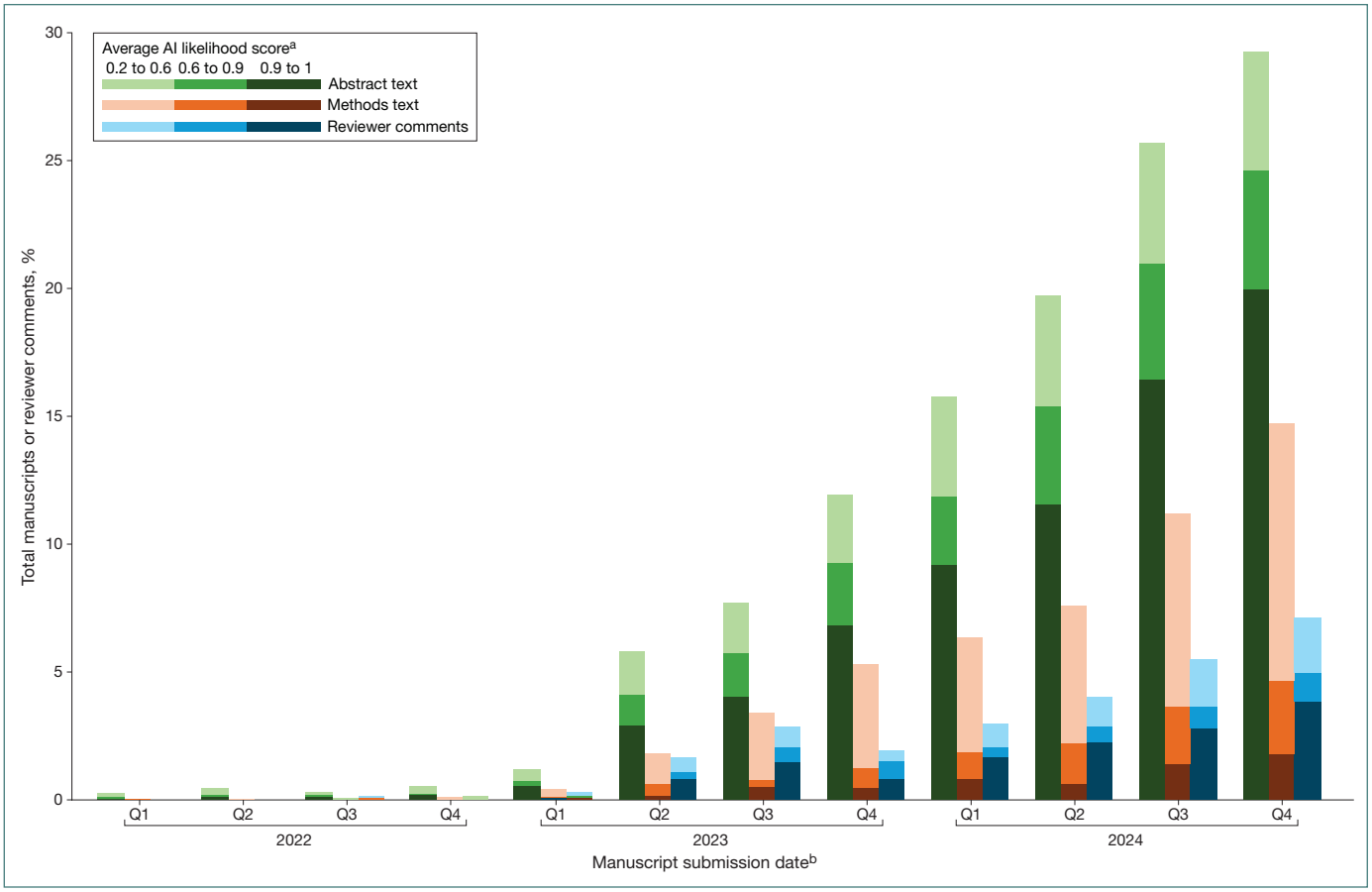
Conclusions AI-generated text can be detected in manuscripts and reviewer comments with virtually no false positives. GenAI usage is increasing rapidly, but journal policies to regulate usage can have a limited impact. Usage varies considerably between journals and the type of text and appears to be associated with different editorial outcomes.

[1]American Association for Cancer Research, Philadelphia, PA, US, 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.

Acknowledgment We thank John Ho for his assistance collecting the data and entering it into the AI detection tool to obtain the AI likelihood scores.

Figure 25-1180. Increase in AI-Generated Text Detected in Submitted Manuscripts and Reviewer Comments Between 2022 and 2024



[a]Average AI likelihood score was calculated over sliding windows approximately 400 words in size.
[b]Manuscripts and reviewer comments were grouped iby submission date based on the date the manuscript passed initial quality control checks.

Comparison of Reasons for Retraction of Biomedical Articles by Women and Men Authors

Ana-Catarina Pinho-Gomes,[1,2] Carinna Hockham,[1] Mark Woodward[1,3]

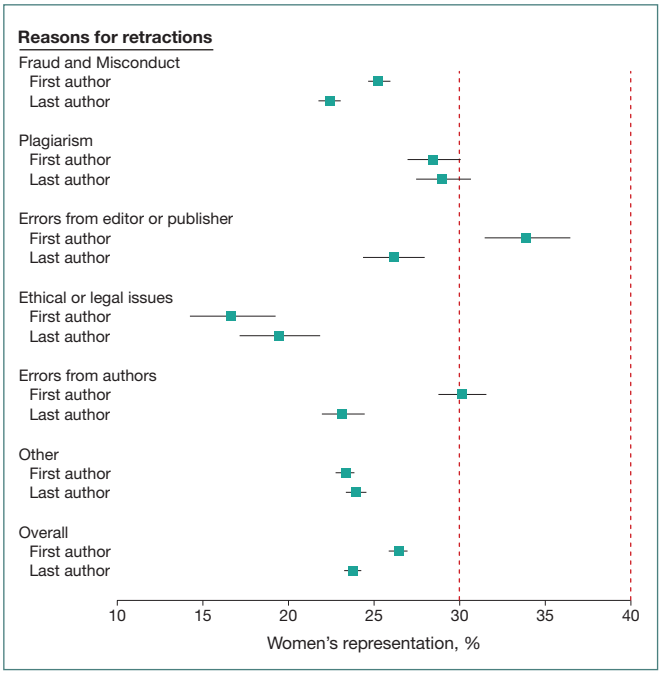
Objective Women are underrepresented among authors of scientific papers. Although the number of retractions has been rising over the past few decades,[1] gender differences among authors of retracted papers remain poorly understood. Therefore, this study investigated gender differences in authorship of retracted articles available on RetractionWatch.

Design RetractionWatch (http://retractiondatabase.org/) provided us a dataset comprising 64,658 articles published between January 1971 and June 2025. This is an update of a previous study.[2] No articles were excluded. We extracted data for the first and last author of each paper and inferred their gender based on their first names using Gender-API software. We accepted gender predictions when the accuracy was estimated to be at least 80%. We estimated the percentage of retractions that could be attributed to women and men first authors and last authors overall and by reason. We used binomial distributions to estimate percentages retracted by women authors, with 95% CIs, and tested whether these percentages were significantly different from 50%. We compared women's representation as authors of retractions vs their representation as authors of articles overall (40% for first authors and 30% for last authors[3]).

Results Among 64,658 articles retracted between 1971 and 2025 that included 64,659 first authors and last authors, women accounted for 26.4% (95% CI, 25.8%-26.9%) of first authors and 23.7% (95% CI, 23.2%-24.2%) of last authors. There were significant differences in gender of authors in retracted articles by reason, and women's representation appeared to be lower than their representation as authors of scientific articles (**Figure 25-0862**). The lowest representation of women was found for ethics and legal issues (16.6% [95% CI, 14.2%-19.2%] for first authors and 19.4% [95% CI, 17.1%-21.8%] for last authors). For first authors, women's representation was the highest for errors caused by editors and publishers (33.8% [95% CI, 31.4%-36.4%]). For last authors, the highest representation of women was for plagiarism (28.9% [95% CI, 27.4%-30.6%]). Most retractions (60.9%) had men as first and last authors.

Conclusions Women's representation among authors of retracted articles seemed to be slightly lower than women's representation as authors of scientific papers overall. Women's underrepresentation as authors of retractions due to fraud and misconduct as well as ethical concerns suggests that gender equality might enhance research integrity and, hence, reduce the negative impact that retractions have on population health and trust in science.

Figure 25-0862. Reasons for Retraction by Proportion of Women First and Last Authors



Error bars indicate 95% CIs.

References

1. Gaudino M, Robinson NB, Audisio K, et al. Trends and characteristics of retracted articles in the biomedical literature, 1971 to 2020. *JAMA Intern Med*. 2021;181(8):1118-1121. doi:10.1001/jamainternmed.2021.1807

2. Pinho-Gomes A-C, Hockham C, Woodward M. Women's representation as authors of retracted papers in the biomedical sciences. *PLoS ONE*. 2023;18(5):e0284403. doi:10.1371/journal.pone.0284403

3. Squazzoni F, Bravo G, Farjam M, et al. Peer review and gender bias: a study on 145 scholarly journals. *Science Advances*. 2021;7(2):eabd0299. doi:10.1126/sciadv.abd0299

[1]The George Institute for Global Health, Imperial College London, London, UK, a.pinho-gomes@imperial.ac.uk; [2]Institute for Global Health, University College London, London, UK; [3]The George Institute for Global Health, University of New South Wales, Sydney, New South Wales, Australia.

Conflict of Interest Disclosures Mark Woodward has been a consultant for Amgen and Freeline. No other disclosures were reported.

Funding/Support Ana-Catarina Pinho-Gomes is funded by an Academic Clinical Lectureship by the National Institute for Health and Care Research (NIHR). Mark Woodward is supported by the National Health and Medical Research Council of Australia grant APP1174120.

Role of Funder/Sponsor The funders had no influence 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 views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Paper Mill Use of Fake Personas to Manipulate the Peer Review Process

Tim Kersjes[1]

Objective The rise of paper mills presents a significant problem for academic publishers. It is suspected that paper mill actors not only pollute the literature by publishing low-quality or even fake articles; they also try and enhance the credibility of their product by citing their own paper mill articles. This reinforces their credibility and attractiveness for paying customers. Papers mills also attempt to subvert the peer review process by suggesting their own actors as peer reviewers. This study is an analysis of how one successful paper mill used journals that rely on author-suggested reviewers to increase their output and credibility through the creation and use of fake authors and reviewers. The objective was to analyze how these fake personas amplified this paper mill's success.

Design Following a research integrity investigation by publisher Springer Nature, 55 articles in 4 of their journals (*Boundary Value Problems*, *Journal of Inequalities and Applications*, *Fixed Point Theory and Applications*, and *Advances in Difference Equations*) were identified to be compromised by a single paper mill, probably operating out of China. These articles were published between September 2013 and August 2020 in journals that asked submitting authors to suggest reviewers. In total, 26 fake personas were identified by confirming with their listed affiliations that these authors were in fact not with these institutions. Network mapping identified the relations between the fake personas and how these aided their publications, as well as citation patterns between articles authored and reviewed by these fake personas.

Results The analysis demonstrated that this paper mill successfully introduced 2 fake personas early on by managing to get these personas as co-authors on published articles. This allowed the paper mill to submit an increasing number of articles (starting with 2 submissions in 2013 to a maximum of 12 submissions in 2016) with new fake personas, with previously published fake personas as suggested reviewers. Because these fake personas infiltrated the peer review process, more fake authors ended up in the published literature, allowing the paper mill to increase its publication output. Of the 26 fake personas, 13 managed to end up as published authors. For these 55 articles, 13 fake personas that are published authors were invited to review for a total of 68 times. In contrast, 13 fake personas without a publication record were only invited 25 times in total. Additionally, 48 of the 55 paper mill articles published by these fake personas cited 2 or more articles by other fake personas in an apparent attempt to further build credibility.

Conclusions Fake personas can serve as both authors and reviewers. Once the paper mill has established a fake persona as credible, it can be used to amplify the credibility of additional fake personas, allowing a paper mill to build a network of fake authors and reviewers to subvert the peer review process. Identity verification and citation analysis may prevent paper mills from gaining success in this way.

[1]Springer Nature, Dordrecht, the Netherlands, tim.kersjes@springernature.com.

Conflict of Interest Disclosures Tim Kersjes is a full-time employee of Springer Nature and owns Springer Nature stock. He also serves on the trustee board of the Committee on Publication Ethics.

Authorship Changes as an Indicator of Research Integrity Concerns in Submissions to Academic Journals

Coromoto Power Febres,[1] Julia Gunn,[1] Laura Wilson[1]

Objective Publishers, industry bodies, and previous literature have identified authorship changes as a potential means of detecting systematic manipulation activity in submissions to academic journals.[1] There is limited empirical evidence on the features associated with systematic manipulation[2] and, consequently, on their accuracy as indicators. This study tested the hypothesis that attempts to change article authorship post submission indicate further concerns.

Design Since March 2023, all requests across 1321 Taylor & Francis journals to change 3 or more authors following initial manuscript submission have been investigated by the Publication Ethics and Integrity team. This threshold was set based on previous experience that 3 or more changes after review was frequently associated with concerning authorship behavior. The request and associated articles were assessed, including adherence to editorial policies and indicators of systematic manipulation.[3] This cross-sectional study analyzed all such assessments conducted in 2024, consisting of 496 articles; 44 articles were excluded due to incomplete assessment. Requests for fewer than 3 changes are assessed by the journal's academic editors, for which data are unavailable. The outcomes measured were the authorship change decision recorded and whether other concerns were noted during assessment. Adherence to editorial policies (eg, ethics and consent declarations, conflict of interest disclosure) and indicators of systematic manipulation (eg, tortured phrases, peer review manipulation) were also recorded; these 2 categories were combined into concerns noted and classified as *yes* or *no*. Odds ratio (OR) analysis tested associations between denied authorship changes and noted concerns. STROBE guidelines informed the study.

Results Of the 452 articles assessed, most requests for 3 or more authorship changes were denied; 367 (81.2%) were denied vs 85 (18.8%) approved. For the 170 articles for which the data were available, the median (range) number of authors was 4 (3-10). A total of 301 requests (66.6%) had concerns and 151 (33.41%) had no concerns. Of the denied requests for authors changes, 292 (79.6%) had concerns and 75 (20.44%) had no concerns. Of approved requests, 76 (89.41%) had no concerns and 9 (10.59%) had concerns. Denied requests were more likely to have concerns (OR, 32.9

[95% CI, 15.9-68.6]) compared with approved requests. The latest decisions per paper per category are shown in **Table 25-1075**.

Conclusions This study suggests that irregularities in authorship changes may indicate research integrity concerns. Limitations include the relatively small sample size and assessment variability across team members/individual assessors. Further analysis is required to establish whether authorship change requests correlate with other systematic manipulation indicators. Further research could also involve analysis by discipline and testing for an association between author demographics and other characteristics, such as co-authorship history and acceptance rate, and a comparison with requests to change 2 or fewer names.

References

1. Abalkina A. Publication and collaboration anomalies in academic papers originating from a paper mill: evidence from a Russia-based paper mill. *Learned Publishing*. 2023;36:689-702. doi:10.1002/leap.1574

2. Byrne JA, Abalkina A, Akinduro-Aje O, et al. A call for research to address the threat of paper mills. *PLoS Biol*. 2024;22(11). doi:10.1371/journal.pbio.3002931

3. Parker L, Boughton S, Bero L, Byrne JA. Paper mill challenges: past, present, and future. *J Clin Epidemiol*. 2024;176. doi:10.1016/j.jclinepi.2024.111549

[1]Taylor & Francis Group, Beeston, UK, coro.powerfebres@tandf.co.uk.

Conflict of Interest Disclosures Coromoto Power Febres reported being a full-time employee of Taylor & Francis and being a member of the STM Research Integrity Committee and United 2 Act Education and Awareness working group. Julia Gunn reported being a full-time employee of Taylor & Francis. Laura Wilson reported being a full-time employee of Taylor & Francis and being a member of the STM Membership committee, United 2 Act Trust markers working group, and ALPSP Policy committee.

Table 25-1075. Latest Decision on Outcome of Submission per Category Analyzed if Data Were Available (n = 434)

Status by request type	Total	With concerns	Without concerns
Rejected			
Pending	49	36	13
Withdrawn	97	87	10
Rejected	106	92	14
Accepted	101	65	36
Subtotal	353	280	73
Accepted			
Pending	7	2	5
Withdrawn	11	2	9
Rejected	16	1	15
Accepted	47	3	44
Subtotal	81	8	73
Total, all requests	434	288	146

Notifying Authors That They Have Cited a Retracted Article and Future Citations of Retracted Articles: The RetractoBot Randomized Controlled Trial

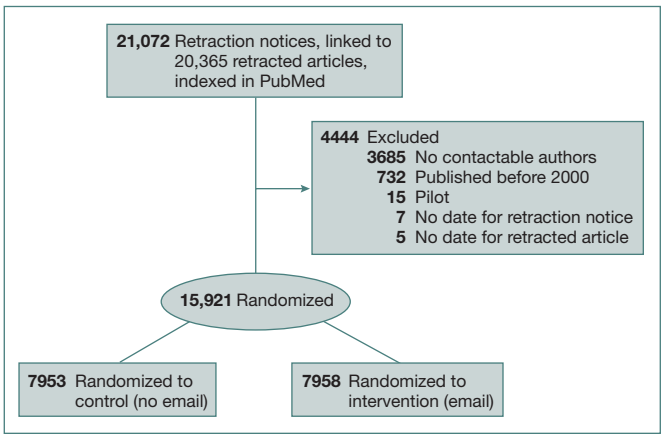
Nicholas J. DeVito,[1] Christine Cunningham,[1] Ben Goldacre[1]

Objective Continued citation of retracted articles is an ongoing issue.[1] RetractoBot examines whether informing authors that they cited a retracted paper reduces future citations to those articles.[2]

Design This was a parallel-group, 2-group, superiority trial. We identified and randomized retracted articles in PubMed published since 2000. The study was unblinded, but primary outcome ascertainment was objective, and randomization and allocation were fully and simultaneously computer automated. Authors with publicly available contact information were notified about citations of retracted articles in the intervention group; no notifications were sent about articles in the control group. The email included a voluntary survey question on knowledge of the retractions and invited additional voluntary feedback. Retraction, contact, and citation data were obtained from PubMed and Scopus. Follow-up started 3 months after the final intervention email was sent to allow in-press manuscripts that could not be affected by our intervention a chance to publish. The primary outcome was the difference in the rate of new citations of retracted articles between groups. We used a negative binomial model adjusted for years since retracted article publication, and our stratification variable, with outliers removed. Data analysis was performed in Stata version 16. We also report the results of a survey question on knowledge of retracted citations, and a sentiment analysis on email feedback. Additional methods details, including power calculations, are available in our prespecified protocol[3] and at retracted.net.

Results We randomized 7958 retracted articles to the intervention group and 7963 to the control group (**Figure 25-0938**). Overall, 246,749 intervention emails were deliverable to authors between January 8 and February 7, 2024, with follow-up extending from May 8, 2024, to May 7, 2025. After removing outliers, there were 7939 citations of articles in the intervention group and 7943 in the control

Figure 25-0938. Flowchart for Study Inclusion and Exclusion



group. The effect of our intervention was not significant (mean citation rate, −0.007; 95% CI, −0.055 to 0.041) and held for all prespecified sensitivity analyses. We received 15,667 survey responses from notified authors with 80.6% (12,631) indicating they were unaware of all retracted citations we informed them about. We received 482 substantive pieces of email feedback, of which 57.7% (278) indicated a positive sentiment about the project (vs neutral, negative, mixed, or other).

Conclusions Our intervention did not significantly decrease citations of retracted articles at 1 year compared with the control group. We plan future secondary analyses to determine whether an effect emerges with increased follow-up. Of the authors who responded, most were unaware of the notified retractions and had a positive view of the intervention, indicating the potential for future interventions in this area. The continued uptake of retraction notifications in citation management software, using sources such as the RetractionWatch database, may already be having an effect in this space.

References

1. Kühberger A, Streit D, Scherndl T. Self-correction in science: the effect of retraction on the frequency of citations. *PLoS One*. 2022;17(12):e0277814.

2. Xu H, Ding Y, Zhang C, Tan BCY. Too official to be effective: an empirical examination of unofficial information channel and continued use of retracted articles. *Res Policy*. 2023;52(7):104815.

3. Cunningham C, DeVito N, Goldacre B. RetractoBot—revised stage 1 registered report protocol. Published online October 31, 2023. doi:10.6084/M9.FIGSHARE.24468391.V1

[1]Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK, nicholas.devito@phc.ox.ac.uk.

Conflict of Interest Disclosures Nicholas J. DeVito reported receiving funding from the Naji Foundation, the Fetzer-Franklin Memorial Fund, the European Commission, and UK Research and Innovation (UKRI) outside the submitted work. Ben Goldacre reported receiving research funding from the Bennett Foundation, the Laura and John Arnold Foundation, the National Health Service (NHS) National Institute for Health Research (NIHR), the NIHR School of Primary Care Research, NHS England, the NIHR Oxford Biomedical Research Centre, the Mohn-Westlake Foundation, NIHR Applied Research Collaboration Oxford and Thames Valley, the Wellcome Trust, the Good Thinking Society, Health Data Research UK, the Health Foundation, the World Health Organization, UKRI Medical Research Council, Asthma UK, the British Lung Foundation, and the Longitudinal Health and Wellbeing strand of the National Core Studies program outside the submitted work; he has previously been a non-executive director at NHS Digital; he also receives personal income from speaking and writing for lay audiences on the misuse of science.

Funding/Support The RetractoBot project was funded by a grant from the Laura and John Arnold Foundation and a grant and in-kind support from Elsevier.

Role of Funder/Sponsor The funders had no role in the study design, data collection (outside of in-kind access to elements of the Scopus API from Elsevier), data analysis, interpretation, decision to submit, or any other facet of the planning, conduct, or dissemination of the presented work.

Acknowledgment We would like to acknowledge the full RetractoBot study team including Karolina Wartolowska, Henry Drysdale, Benjamin Feakins, Helen J. Curtis, Francis Irving, Anna Powell-Smith, and Seb Bacon.

How a Questionable Research Network Manipulated Scholarly Publishing

Leslie D. McIntosh,[1] Hélène Draux,[1] Elizabeth Smee,[1] Cynthia Hudson Vitale[2]

Objective To understand the broad impact of one questionable network on publishing, including the publishers, countries, and institutions involved and affected by the network's activity.

Design We examined the Pharmakon Neuroscience Research Network (PNN), a questionable entity listed in funding statements, acknowledgments, and author affiliations in articles published between 2019 and 2023. We selected the PNN because it was the first potential broad-scale case of scientific manipulation identified by one of the authors (L.D.M), and analyzing the PNN has helped the development of patterns to discover other unusual collaboration networks. All documents containing "Pharmakon Neuroscience" anywhere in the title, abstract, author affiliations, and full text were selected using Dimensions.[1] We excluded retraction and correction notices; articles for which a peer reviewer identified their organization as PNN; and articles with more than 25 authors. We inspected the metadata associated with the articles, notably the organizations related to the articles. We situated these practices within the Taxonomy of Scientific Manipulation,[2] systematically identifying (1) who was involved (authors, institutions, and publishers), (2) where mechanisms were used to manipulate science (journals, countries, and institutions), and (3) how the science was manipulated (methods for deceiving scholarly communication).

Results Pharmakon Neuroscience was referenced in 140 published articles as of January 2025. After applying exclusion criteria, 123 articles remained. These articles included more than 6000 citations (citation ratio of 51 citations per article). The articles had 361 unique authors, including 29 without author identifiers (eg, ORCiD, Dimensions). The PNN involved 56 journals across 12 publishers, with Springer Nature (n = 29), Elsevier (n = 26), and Bentham Science Publishers (n = 25) having the most articles involved out of the total (n = 123). The involved authors were affiliated with 40 countries and 232 organizations. Of these organizations, 212 had research identifiers (eg, Research Organization Registry, Global Research Identifier Database), and 20 did not. Three of the organizations appeared to be personal homes, even though they were claimed as businesses or educational institutions. Additionally, one publication[3] highlighted the PNN as an

affiliation among the "top 25 most predictive and impactful authors."

Conclusions This investigation leveraged publication databases and advanced forensic scientometric techniques to highlight critical research integrity and security vulnerabilities in a questionable research network. We use the term "questionable network" because this co-authorship network appears to have grown too rapidly to have developed naturally from scholarly affiliations over time. The impact of the PNN's work might be legitimized through future studies and become seen as normative behavior. Although the most prolific PNN author and organization stem from one country, there is a global impact not confined to high- or low-income countries. While one individual appears central to the network, a single orchestrator does not imply sole responsibility. Instead, this case underscores the complexity of manipulating the scientific ecosystem, where questionable actors exploit the scholarly communication system through nonverifiable funders and organizations and artificially inflated citation metrics.

References

1. Hook DW, Porter SJ, Herzog C. Dimensions: building context for search and evaluation. *Front Res Metr Anal*. 2018;3:23. doi:10.3389/frma.2018.00023

2. McIntosh LD, White W, Vitale CH. Unveiling deception: establishing a taxonomic framework for disinformation within scientific discourse. *arXiv*. Preprint posted online November 19, 2023. doi:10.48550/arxiv.2311.11344

3. Sab MC, Ahmed KM. Tracing Bangladesh's scientific footprints: a scientometric expedition into inflammation pharmacology. *Inform Res Comm*. 2024;1(1):9-21. doi:10.5530/irc.1.1.4

[1]Digital Science, London, UK, l.mcintosh@digital-science.com; [2]Johns Hopkins University, Baltimore, MD, US.

Conflict of Interest Disclosures Leslie D. McIntosh, Hélène Draux, and Elizabeth Smee are employees of Digital Science, a research technology company.

Diversity and the Research Environment

An Analysis of Equity, Diversity, and Inclusion Concerns From JAMA Network Peer Reviewers

Michael O. Mensah,[1] Anand R. Habib,[2] Jacob Kendall-Taylor,[3] Mya Roberson,[4] Kanade Shinkai,[4] Annette Flanagin,[3] Preeti Malani[3]

Objective In March 2023, the JAMA Network invited peer reviewers to check a box when concerned about elements of equity, diversity, and inclusion (EDI) in manuscripts, and to explain such concerns in confidential comments to the editor. We examined peer reviewers' use of the EDI concerns checkbox and categorized concerns described in confidential comments.

Design In this cross-sectional, mixed-methods analysis, we initially examined all reviewed manuscripts for 13 JAMA Network journals between March 28, 2023, and December 31, 2024, and determined counts and percentages of reviews that utilized an EDI concerns checkbox. As a case study, we then used rigorous and accelerated data reduction (RADaR) to analyze deidentified, confidential reviewer comments for *JAMA Dermatology* manuscripts. RADaR develops all-inclusive data tables, then iteratively reduces those tables until only research question–relevant data remain.[1] Two former *JAMA* editorial fellows independently classified confidential comments as either relevant or irrelevant to EDI and then inductively coded comments. The coders agreed on which code(s) best described each comment and then distilled prominent categories among all comments. This study was deemed exempt by Yale School of Medicine's institutional review board and follows STROBE and SRQR guidelines.

Results Among 39,151 reviews across all JAMA Network journals, 5.3% (2075 of 39,151) utilized the EDI concerns checkbox, including 5.5% (1897 of 34,542) of research manuscript reviews. *JAMA Otolaryngology* reviews utilized the EDI checkbox least often (2.63% [29 of 1104] overall; 2.7% [26 of 948] of research manuscripts). *JAMA Pediatrics* reviews utilized the EDI checkbox most often (8.14% [146 of 1793]; 8.0% [126 of 1583] of research manuscripts). *JAMA Dermatology* reviews utilized the EDI checkbox at the median frequency among all JAMA Network journals (4.84% [91 of 1879]; 5.4% [78 of 1439] of research manuscripts) (**Table 25-1090**). Among *JAMA Dermatology* reviews utilizing the EDI checkbox, 34.1% (31 of 91) had confidential comments relevant to EDI, 50.5% (46 of 91) were irrelevant to EDI, and 15.4% (14 of 91) had no comment. Regarding EDI-relevant confidential comments, categories of responses included lack of clarity regarding race and ethnicity categories (48.4% [15]), statistical or methodological concerns (16.1% [5]), skin tone representation (9.7% [3]), ambiguity in language around demographic identifiers (9.7% [3]), and inappropriate use of gender ascribed by observers (6.5% [2]).

Conclusions Overall, the EDI concerns checkbox was infrequently used across the JAMA Network journals. Among *JAMA Dermatology* reviews using the EDI concerns checkbox, less than half had confidential comments relevant to EDI, perhaps because instructions to reviewers did not define EDI concerns. Prominent categories among EDI-relevant comments included concerns about proper usage of race and ethnicity categories, absence of photos depicting dermatologic conditions in darker skin tones, and use of inclusive language regarding sex and gender, consistent with equity issues previously noted by the American Academy of Dermatology.[2] Future work could explore ways of systematically capturing EDI concerns from reviewers and their eventual influence on manuscripts accepted for publication.

References

1. Watkins DC. Rapid and rigorous qualitative data analysis: the "RADaR" technique for applied research. *Int J Qualitative*

Table 25-1090. Proportion of Manuscript Peer Reviews Across JAMA Network Journals With a Reviewer-Identified Equity, Diversity, and Inclusion (EDI) Concern

Journal	Total No. of manuscript peer reviews	Manuscript peer reviews utilizing the EDI checkbox	
		No.	Percentage
JAMA	8048	370	4.60
JAMA Cardiology	1691	71	4.20
JAMA Dermatology	1879	91	4.84
JAMA Health Forum	1352	68	5.03
JAMA Internal Medicine	1645	74	4.50
JAMA Network Open	12,766	829	6.49
JAMA Neurology	1896	94	4.96
JAMA Oncology	1521	78	5.13
JAMA Ophthalmology	1064	29	2.73
JAMA Otolaryngology–Head & Neck Surgery	1104	29	2.63
JAMA Pediatrics	1793	146	8.14
JAMA Psychiatry	1406	98	6.97
JAMA Surgery	2986	98	3.28
Median (IQR)	1691 (1379-2441)	69.5 (91-122)	4.84 (3.74-5.81)
Overall	39,151	2075	5.30

Methods. 2017;16(1):1609406917712131. doi:10.1177/1609406917712131

2. American Academy of Dermatology Association. Diversity and the academy. Accessed June 8, 2025. https://www.aad.org/member/career/diversity

[1]Department of Psychiatry, Yale School of Medicine, New Haven, CT, US, michael.mensah@yale.edu; [2]National Clinician Scholars Program, Yale School of Medicine, New Haven, CT, US; [3]*JAMA* and the JAMA Network, Chicago, IL, US; [4]*JAMA Dermatology*, Chicago, IL, US.

Conflict of Interest Disclosures Michael O. Mensah is co-editor of the Race and Mental Health Equity column in *Psychiatric Services*, unrelated to this work. Mya Roberson receives consulting fees from the National Committee for Quality Assurance outside the scope of the submitted work. Annette Flanagin is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Assessment of an Intervention to Equalize the Proportion of Funded Grant Applications for Underrepresented Groups at the Canadian Institutes of Health Research

Anne Lasinsky,[1] James Wrightson,[2] Matthew Hogel,[3] Alannah Brown,[3] Adrian Mota,[3] Karim M. Khan,[1,2,4] Clare L. Ardern[5,6]

Objective A small number of Canadian research funders have implemented interventions to address known biases in grant peer review. Equalization, which aims to match the proportion of funded grant applications to the proportion of submitted applications from specific underrepresented groups, is one such intervention. In 2016, the Canadian Institutes of Health Research (CIHR) implemented equalization for early career researcher (ECR) principal applicants in its Project Grant Competition. In 2021, equalization expanded to female principal applicants and French-language applicants. The objective of this study was to describe the outcome of equalization in 2022.

Design This was a retrospective analysis of the number of funded grants in the spring 2022 CIHR Project Grant Competition. The equalization intervention was applied to applications submitted by ECR applicants, female principal applicants, and applications in French. To apply the intervention, first, scores for all applications submitted to the Project Grant Competition, which were reviewed by 59 committees, were converted to a percentage rank to account for scoring differences across committees. CIHR funded applications in rank order from the top percent rank, as far down as the competition budget allowed, and intervened to ensure that the next-ranked applications from each underrepresented applicant group were funded. Equalization matched the proportion of applications funded to the proportion of applications submitted by each group. No grants were defunded. Any grants that were equalized were added to the pool of funded grants—they did not displace another applicant's score-win. This study descriptively analyzed routinely collected data from CIHR. The main outcome was the number of grant applications funded for each underrepresented applicant group, with and without equalization. The secondary outcome was the proportion of funded applications for each underrepresented applicant group, with and without equalization.

Results There were 2095 applications submitted to the spring 2022 Project Grant Competition. Before equalization, 370 applications (17.7%) were funded. After equalization, 405 applications (19.3%) were funded with a total of CAD$325 million. ECR principal applicants submitted 580 applications (27.7%), female principal applicants submitted 774 applications (36.9%), and 25 applications (1.2%) were submitted in French). After equalization, 21 additional applications from ECR principal applicants and 22 applications from female principal applicants were funded (**Table 25-1092**). The funding success rates increased from 16.9% to 20.7% for ECR principal applicants and 17.5% to 20.3% for female principal applicants. One additional French-language application was funded with equalization; the success rate for French-language applicants increased from 17.4% to 21.7%.

Conclusions In the spring 2022 CIHR Project Grant Competition, equalization increased the number of health research grants awarded and the funding success rate for ECR

Table 25-1092. Grant Success Rates With and Without Equalization in 2022 (370 Applications Were Funded Before Equalization; 405 Applications Were Funded After Equalization)

Applicant	Grants awarded (No.)			Success rate (%)	
	Before equalization	Added	Total	Before equalization	After equalization
ECR principal applicants					
Applications	92	21	113	15.9	19.5
NPIs	91	21	112	16.9	20.7
Female principal applicants					
Applications	128	22	150	16.5	19.4
NPIs	127	22	149	17.5	20.3
French applications					
Applications	4	1	5	16.0	20.0
NPIs	4	1	5	17.4	21.7

Abbreviations: ECR, early career researcher; NPI, nominated principal investigator.

and female principal applicants, and for applications submitted in French.

[1]School of Kinesiology, The University of British Columbia, Vancouver, Canada; [2]Department of Family Practice, The University of British Columbia, Vancouver, Canada; [3]Canadian Institutes of Health Research, Ottawa, Canada; [4]Canadian Institutes of Health Research-Institute of Musculoskeletal Health and Arthritis, Vancouver, Canada; [5]Department of Physical Therapy, The University of British Columbia, Vancouver, Canada, clare.ardern@ubc.ca; [6]Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Australia.

Conflict of Interest Disclosures Matthew Hogel is Deputy Director, Funding Analytics at the Canadian Institutes of Health Research (CIHR). Alannah Brown is Senior Advisor to the Associate Vice-President at CIHR. Adrian Mota is Acting Vice President, Research—Programs at CIHR. Karim M. Khan is Scientific Director for CIHR's Institute of Musculoskeletal Health and Arthritis (2017-2025). No other conflicts were reported.

Funding/Support This work was supported by a CIHR Research Operating Grant (Scientific Directors) held by Karim M. Khan. CIHR's Funding Analytics coordinated data management and analysis as part of its mandate to foster and deliver high quality peer review for health research in Canada.

Role of the Funder/Sponsor CIHR did not participate in preparing the abstract, nor in the decision to submit the abstract for presentation.

Extracting Research Environment Indicators From the UK Research Excellence Framework 2021 Statements

Noémie Aubert Bonn,[1,2] Lukas Hughes-Noehrer[1]

Objective The impact that research environments have on the quality of research has gained heightened visibility in the past few years. The Research Excellence Framework (REF), a UK-wide funding exercise that assesses higher education institutions (HEIs) on the excellence of their research, includes research environment as part of the elements it assesses. Recently, the weight of research environment was increased to one-fourth of the total score, bringing it closer to the other elements of research outputs and impacts. However, measuring positive research environments is challenging and clear indicators are still hard to delineate.

Design To identify indicators that may be associated with positive research environments, we conducted a qualitative analysis of environment statements submitted to REF 2021.[1] We sampled 34 four-star statements drawn from different units of assessment (ie, academic disciplines) available openly on the REF website. Four-star scores are given to statements that describe "an environment that is conducive to producing research of world-leading quality and enabling outstanding impact, in terms of its vitality and sustainability." Using these statements, 1 researcher conducted a mix of inductive (first 17 statements in NVivo 12) and deductive (last 17 statements) thematic analysis to extract the potential indicators of positive research environments. These indicators were organized in core themes, discussed with the research team, and compared with existing recommendations, policies, and literature on healthy research environments to build an exhaustive overview of possible indicators.

Results The analysis identified a vast array of elements, which were used to showcase positive research environments. We organized the elements extracted in indicators providing details on people, such as diversity metrics, staff and student characteristics, and measures to improve equity, diversity, and inclusion; indicators demonstrating positive research cultures, such as responsible leadership, collaborative culture, research integrity, and open research; and indicators demonstrating positive research environments, such as capacity-building measures, staff and financial support, employment security and progression, and responsible assessment practices (**Table 25-0889**). The breadth of indicators identified showcased differences and similarities in how different institutions and units of assessment qualified their research environments.

Conclusions Our analysis provided insights on the breadth of elements that can come into play in creating a positive research environment. Despite being drawn from the UK landscape, our findings could be relevant to a broad diversity of research settings. For example, our findings could help

support more comprehensive research assessment or they could help inspire concrete actions to improve research environments in HEIs.

Table 25-0889. Core Themes and Indicator Groups Extracted

Indicators providing details on people	
Staff and student snapshots	Staffing snapshot, student snapshot
EDI snapshot	EDI snapshot, EDI in REF preparation, EDI awards, capture and reporting of EDI elements
Measures to improve EDI	Staff and committees dedicated to EDI, training in EDI and unconscious bias, EDI policies and action plans, inclusive recruitment, inclusive promotion, support measures for underrepresented groups, parental leave and family support, pay gap measures, grassroots initiatives and advocacy, visibility and accessibility of EDI measures, measures against bullying and harassment
Indicators demonstrating positive research culture	
Responsible leadership	Leadership and management awards, collaborative leadership
Collaborative culture	Culture building, collaborations, peer review
Diversity of research contributors	Early career contributors, diverse contributing sectors
Responsible research practices: research integrity	Commitment to research integrity, oversight and governance, research integrity training, retractions and corrections
Responsible research practices: open science	Open access; open project governance; project policies on open science; open materials, codes, and methods; open review; open data; open-source software; open training materials; preprinting; licensing; data preservation and management policies; bottom-up involvement and advocacy; open science awards
Responsible research practices: reproducibility	Preregistration, reproducibility initiatives
Indicators demonstrating positive research environments	
Capacity building	Formal support for capacity building, training opportunities, customized training opportunities, interuniversity and faculty training programs
Support for ECRs	ECR support and integration, doctoral student support and integration, evidence of success stories from ECRs, scholarships and financial support
Support for technical staff	Formal commitment to support staff, support for technical staff, involvement of technical staff in department activities
Financial support	Bench fee and seed funding, research funding support
Employment security and career advancement	Employment security and career advancement snapshot, support for researchers to undertake careers outside academia, ECRs employed in academic positions, ECRs employed in nonacademic positions, retention and career continuity strategies, transition funds, support for career advancement
Balanced workloads and flexible working solutions	Efforts to promote balance in tasks, research leave and secondment opportunities, parental leave and family support, flexible working solutions
Mental health and well-being support	Well-being surveys, healthy workplace awards, dedicated resources to mental health and well-being
Responsible research assessment practices	Commitment to responsible research assessment initiatives, responsible recruitment practices, responsible assessment practices

Abbreviations: ECR, early career researcher; EDI, equity, diversity, and inclusion; REF, Research Excellence Framework.

Reference

1. Research Excellence Framework. Results and submissions. May 12, 2022. Accessed June 24, 2025. https://results2021.ref.ac.uk/

[1]The University of Manchester, Department of Computer Science, Faculty of Science and Engineering, Manchester, UK, noemie.aubertbonn@manchester.ac.uk; [2]Hasselt University, Department of Healthcare & Ethics, Faculty of Medicine, Hasselt, Belgium.

Conflict of Interest Disclosures Noémie Aubert Bonn was employed as a senior policy advisor with Research England from 2022 to 2023.

Funding/Support The conceptualization, design, and part of the analysis was carried out and supported during Noémie Aubert Bonn's employment at Research England. The analysis software subscription was provided by Hasselt University, Belgium.

Role of the Funder/Sponsor The conclusions of the study are independent from Research England.

Acknowledgment We wish to thank all the members of the People, Culture and Environment team at Research England who provided feedback and discussion throughout this project, particularly Catriona Firth, Claire Fraser, Steven Hill, Myles Furr, Marie-Helene Nienaltowski, Duncan Shermer, and Nicole Dixon. We also wish to thank everyone within UK Research and Innovation and The University of Manchester who provided feedback and insights on this project.

Research Misconduct and Integrity

Retractions and Democracy Index Scores Across 167 Countries

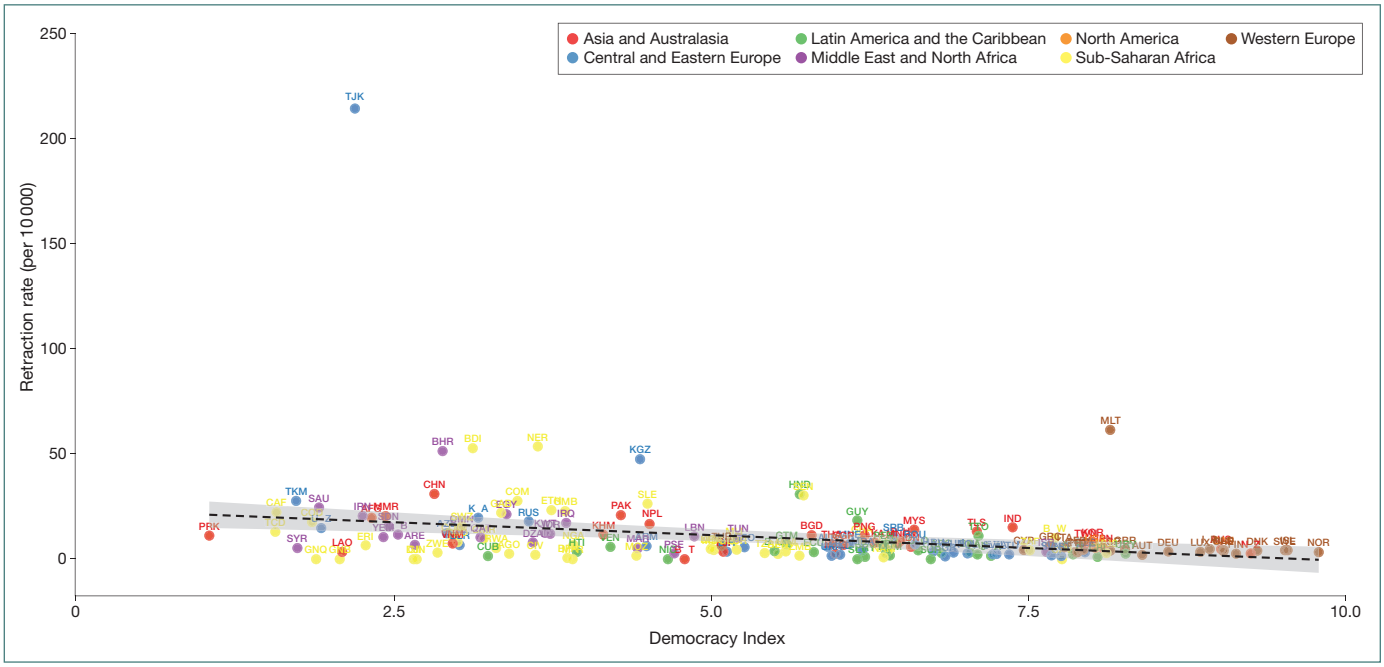
Ahmad Sofi-Mahmudi,[1] Hesam Salmabadi[2]

Objective To evaluate the association between countries' democratic status and scientific article retraction rates while exploring potential factors influencing this association.

Design This retrospective cohort study analyzed publication and retraction data from 2006 to 2023. Data sources included the Retraction Watch Database for retraction data (n = 64,264 retractions), the Economist Intelligence Unit's Democracy Index for democratic status, and SCImago for country-specific publication outputs (n = 59,385,751 publications). Our primary outcome was retraction rate (retractions per 10,000 publications) rather than absolute retraction counts, to account for varying publication volumes across countries. Using bayesian hierarchical negative binomial regression models with multiple imputations for missing data, we examined the association between the democracy index and article retractions while controlling for key confounders, including gross domestic product (GDP) per capita, research investment as percentage of GDP, English language proficiency scores, corruption control, government effectiveness, regulatory quality, rule of law, press freedom, and international collaboration rates, across 167 countries.

Results The analysis of 167 countries demonstrated a median democracy index score of 5.71 (range, 1.05-9.81) and a median scientific publication retraction rate of 5.3 per

Figure 25-1106. Regional Patterns in the Association Between Democracy Index Scores and Scientific Publication Retraction Rates (2006-2023)



10,000 publications (range, 0-214.8). Higher democracy scores were associated with lower retraction rates (adjusted coefficient, −0.46; 95% credible interval [CrI], −0.59 to −0.32) after controlling for confounders (**Figure 25-1106**). Among institutional factors, rule of law showed a positive association (0.40; 95% CrI, 0.17-0.63), whereas corruption control demonstrated a negative association (−0.54; 95% CrI, −0.72 to −0.36) with retraction rates. The association between democracy and retractions remained stable across periods (interaction coefficient, −0.03; 95% CrI, −0.46 to 0.39), although press freedom effects varied significantly between the 2006 to 2012 and 2013 to 2023 periods (interaction coefficient, −1.40; 95% CrI, −1.77 to −1.03). GDP per capita (0.09; 95% CrI, 0.01-0.17) and government effectiveness (0.21; 95% CrI, 0.01-0.41) were positively associated with retraction rates, suggesting better detection and correction mechanisms in wealthier, more effectively governed countries.

Conclusions Democratic status shows a negative association with scientific article retraction rates, although this association is complex and mediated by institutional factors. The findings suggest that while democratic societies may have lower rates of problematic research requiring retraction, they also demonstrate stronger institutional capacity for detecting and addressing research misconduct.

[1]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada, a.sofimahmudi@gmail.com; [2]Department of Environmental Sciences, University of Quebec in Trois-Rivières, Trois-Rivières, Quebec, Canada.

Conflict of Interest Disclosures Ahmad Sofi-Mahmudi is an employee of Cytel Canada Health Inc.

Additional Information We used Claude, 3.5 Sonnet (claude-3-5-sonnet-20241022) through its website for refining the analysis codes. Also, it was used as a grammar and flow checker of the abstract.

Characterizing Problematic Images in Retracted Scientific Articles

João Phillipe Cardenuto,[1] Daniel Moreira,[2] Anderson Rocha[1]

Objective To quantitatively analyze the types, contexts, and manipulation methods of problematic images that lead to retractions of scientific articles.

Design This cross-sectional study analyzed retracted articles flagged for problematic image manipulation (eg, image duplication) in the Retraction Watch Database[1] (56,716 entries as of October 4, 2024). We focused on entries containing the term *image* in the retraction reason (8002 entries) and further refined the dataset to those discussed on PubPeer[2] (2078 after duplicate removal) to gain more detailed insights into the image problems. Data extracted included figure types (eg, microscopy, gel blot), the context of image misuse (eg, within-article, between-article), and the type of manipulation (eg, duplication, splicing).

Results The Table presents our results (**Table 25-0965**). Gel blots (eg, Western blots) were the most frequently cited image type in problematic retractions, appearing in 1074 articles (51.68%). Between-article image reuse, where an image and its associated data are duplicated across different publications, was the most common context of misuse, identified in 1241 cases (59.72%). Notably, 982 retractions (47.28%) were attributed to paper mills. Image duplication was the predominant cause of retraction, accounting for 1827 cases (87.92%). Only 1 retraction was attributed to computer- or artificial intelligence–generated manipulation. While our analysis did not filter by the biomedical area, most problematic images originated from the biomedical domain.

Table 25-0965. Characterizing Problematic Images in Retracted Articles

Characteristic	Retracted articles, No. (%) (N = 2078)[a]
Most prevalent types of images	
Gel blots	1074 (51.68)
Transwell assay[b]	569 (27.38)
Microscopy imagery	509 (24.50)
Fluorescence-activated cell sorting	355 (17.08)
Fluorescent microscopy	248 (11.94)
Graphs[c]	186 (8.95)
Wound healing assay[d]	181 (8.71)
Colony formation assay[e]	176 (8.47)
Exposed organ (tumor)[f]	171 (8.23)
Electron microscopy	81 (3.90)
Other	131 (6.30)
Problem context	
Between articles	1241 (59.72)
Paper mill	982 (47.26)
Within figures	810 (38.98)
Within images	402 (19.35)
Between figures	344 (16.55)
Problem type	
Reuse	1827 (87.92)
Splicing	162 (7.80)
Image edit, removal, or obscuring	79 (3.80)
Computer- or artificial intelligence–generated	1 (0.05)
Other	346 (16.65)

[a]A single retracted article may contain multiple types of problematic images or contexts, so the percentages do not sum to 100.

[b]Cell migration and invasion studies.

[c]Linear plots, bar plots, or scatter plots that are manipulated or duplicated.

[d]Studies depicting 2-dimensional cell migration throughout an artificial gap.

[e]In vitro cell colony growth studies.

[f]Exposed organs, such as mouse brain or lung slices, predominantly from mice.

Conclusions This study highlights the prevalence of gel blot images and between-article image duplication in retracted articles, indicating a potential benefit from specialized tools to detect such issues. During our analysis, we noticed a frequent lack of detailed and standardized information in retraction notices, which hinders efforts to understand and prevent the presented problems. While PubPeer data offer valuable insights when the retraction notices fail to do so, PubPeer posts are not official documents and may exhibit biases from their authors, which could result in speculative claims about an article. Because of that, to facilitate research and improve the integrity of the scientific record, future research should focus on discussing and developing better guidelines for comprehensive retraction notices that may even support computer-aided solutions.

References

1. The Retraction Watch Database. Accessed January 31, 2025. http://retractiondatabase.org/

2. PubPeer. Accessed January 31, 2025. https://pubpeer.com

[1]Artificial Intelligence Lab, Recod.ai, Institute of Computing, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil, phillipe.cardenuto@ic.unicamp.br; [2]Department of Computer Science, Loyola University Chicago, Chicago, IL, US.

Conflict of Interest Disclosures None reported.

Funding/Support This study was supported by the National Council for Scientific and Technological Development–CNPq (grant No. 442229/2024-0) and by the São Paulo Research Foundation–FAPESP (grant No. 2023/12865-8).

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.

Misidentification of Scanning Electron Microscope Instruments in the Peer-Reviewed Materials Science and Engineering Literature

Reese A. K. Richardson,[1,2] Jeonghyun Moon,[1] Spencer S. Hong,[1,3] Luis A. Nunes Amaral[1,2,4,5,6]

Objective Materials science and engineering (MSE) research has, for the most part, escaped the doubts raised about the reliability of scientific literature by recent large-scale replication studies in psychology and cancer biology. However, users on postpublication peer review sites have recently identified dozens of articles where the make and model of the scanning electron microscope (SEM) listed in the text does not match the instrument's metadata visible in the images in the published article.

Design To systematically investigate this potential risk to the MSE literature, we developed a semiautomated approach using optical character recognition to scan published figures for instrumental metadata banners and check the information contained in these banners on instrument manufacturer and model against the SEM instrument identified in the text. We validated our approach by evaluating model performance on known instances of SEM misidentification and manual reevaluation of novel instances of SEM misidentification returned by the pipeline. This analysis was performed in January 2024.

Results Starting from a set of 1,067,102 articles published since 2010 (as indexed by OpenAlex, downloaded on February 27, 2023) in 50 MSE journals (selected according to the publisher's permissiveness of text and data mining and to represent a broad range of subdisciplines and impact factors), we identified 11,314 articles for which SEM make and model could be identified in an image's metadata. For 2400 of those articles (21.2%), the image metadata did not match the SEM manufacturer or model listed in the text, and for another 2799 (24.7%), at least some of the instruments used in the study were not reported. We found that articles with SEM misidentification were more likely to have existing PubPeer

comments than articles without (43 of 2400 [1.8%] vs 984 of 171,646 [0.5%]; $P = 9.98 \times 10^{-15}$) and that electrochemistry articles that featured SEM misidentification, compared with those that did not have SEM misidentification, were more likely to make a crucial error in their determination of optical band gap (41 of 154 [26.6%] vs 100 of 751 [13.3%]; $P = 3.35 \times 10^{-5}$).

Conclusions Our results suggest that SEM misidentification can be a valuable indicator of irreproducibility in the results of extant scientific articles. We recommend that peer reviewers, before and after publication, remain attentive to the accurate reporting of instruments used in scientific articles. Our selection criteria for journals and the high false-negative rate of our pipeline limit the extent to which our findings are representative of the MSE literature at large. Unexplained textual similarities (eg, the same unlikely typo appearing in dozens of articles) common to many of these articles suggest the involvement of paper mills, organizations that mass produce, sell authorship on, and publish fraudulent scientific manuscripts at scale.

[1]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL US, richardsonr43@gmail.com; [2]Department of Molecular Biosciences, Northwestern University, Evanston, IL US; [3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL US; [4]Department of Physics and Astronomy, Northwestern University, Evanston, IL, US; [5]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, US; [6]NSF-Simons National Institute for Theory and Mathematics in Biology, Chicago, IL, US.

Conflict of Interest Disclosures None reported.

Funding/Support Reese A. K. Richardson was supported in part by the National Institutes of Health Training grant T32GM008449 through Northwestern University's Biotechnology Training Program; Reese A. K. Richardson gratefully acknowledges funding from the Dr. John N. Nicholson fellowship from Northwestern University and Moderna Inc. Spencer S. Hong gratefully acknowledges support from the Ryan Fellowship and the International Institute for Nanotechnology at Northwestern University; Luis A. Nunes Amaral gratefully acknowledges funding from SCISIPBIO: a data-science approach to evaluating the likelihood of fraud and error in published studies (grant No. 1956338).

Role of the Funder/Sponsor The funders had no role in the design or execution of this study.

Indicators of Small-Scale and Large-Scale Citation Concentration Patterns

Iakovos Evdaimon,[1] John P. A. Ioannidis,[2] Giannis Nikolentzos,[3] Michail Chatzianastasis,[1] George Panagopoulos,[4] Michalis Vazirgiannis[1,5]

Objective Citation counts are widely used to evaluate research influence.[1] The h-index, while popular, does not account for how citations are distributed across articles or authors. This is critical when citations are strategically concentrated to boost perceived influence. We studied extreme citation concentration patterns that may distort bibliometric evaluations and merit further scrutiny.[2]

Design We analyzed Scopus data up to March 2023, focusing on 1,496,680 authors with more than 5 full articles and at least 1000 citations. We used Scopus-calculated h-indices and citation counts and introduced 3 indicators of extreme citation concentration: C/h^2 ratio, $A_{50\%C}$, and A_{50}. For C/h^2 ratio, a small value (first percentile, 2.45) indicates that the citations of the author may have been strategically placed to increase the h-index. This metric normalizes citation volume by what would be expected if the h-index reflected broad citation distribution. $A_{50\%C}$ measured the number of distinct citing authors contributing at least 50% of an author's total citations. Extremely low values (first percentile, 5) suggest reliance on few citing sources, possibly from self-citation or closed groups. A_{50} measured the number of coauthors with whom an author has published more than 50 articles. Extremely high values (99th percentile, more than 7) may indicate large-scale collaborations and dense citation clusters. Thresholds were based on the 1% extremes of each metric's distribution. Authors in physics and astronomy were excluded from A_{50} and $A_{50\%C}$ analyses due to the influence of nuclear and particle physics. Field definitions followed the Science-Metrix classification.

Results In the C/h^2 analysis, 14,967 authors (1%) fell below 2.45. These authors had a median of 63 articles, all with h-indices exceeding 21. Fields such as chemistry and earth and environmental sciences showed 1.5-fold higher representation of authors with extremely low C/h^2 scores (**Table 25-0974**). In the $A_{50\%C}$ analysis, 11,277 authors (1%) had 5 or fewer citing authors accounting for 50% of their citations. The median number of articles was 104, and most authors had an h-index greater than 20. Fields with the highest enrichment included chemistry and mathematics and statistics. In the A_{50} analysis, 12,015 authors (1%) had more than 7 highly recurrent coauthors (50 or more shared publications). These authors were exceptionally prolific (median, 596 articles) and had very high h-indices (median, 56; 12.8% with an h-index greater than 100). A total of 73.3% were in clinical medicine. Co-occurrence analysis revealed a strong association between low C/h^2 and low $A_{50\%C}$ values (odds ratio, 6.4; 95% CI, 5.9-6.9). There was modest coexistence for extremely low $A_{50\%C}$ and extremely high A_{50} values (odds ratio, 1.5; 95% CI, 1.3-1.7). Extremely low C/h^2 very rarely co-occurred with extremely high A_{50} values (odds ratio, 0.09; 95% CI, 0.05-0.16).

Conclusions Extreme patterns of citation concentration can substantially distort citation-based metrics, especially the h-index. While not implying misconduct, these metrics highlight outliers where citation behavior diverges from norms. Such metrics may serve as red flags for further analysis.

References

1. Bornmann L. How are excellent (highly cited) papers defined in bibliometrics? a quantitative analysis of the literature. *Res Eval.* 2014;23(2):166-173. doi:10.1093/reseval/rvu002

Table 25-0974. Authors With Extreme Metrics: Distribution Across the 22 Main Fields of Science

Field of study	Authors, No. (%)			
	Total (N = 1,496,680)	Lowest 1% of C/h^2 (n = 14,967)	Lowest 1% of $A_{50\%C}$[a] (n = 11,277)	Upper 1% of A_{50}[a] (n = 12,015)
Agriculture, fisheries, and forestry	35,839 (2.4)	859 (5.7)	247 (2.2)	172 (1.4)
Biology	64,737 (4.3)	897 (6.0)	489 (4.3)	90 (0.8)
Biomedical research	189,457 (12.7)	1824 (12.2)	537 (4.8)	1085 (9.0)
Built environment and design	58,067 (0.2)	37 (0.3)	50 (0.4)	2 (<0.1)
Chemistry	20,776 (6.1)	2065 (13.8)	1745 (15.5)	369 (3.1)
Clinical medicine	579,366 (38.7)	4370 (29.2)	1927 (17.1)	8806 (73.3)
Communication and textual studies	1617 (0.1)	4 (<0.1)	29 (0.3)	0
Earth and environmental sciences	58,067 (3.9)	888 (5.9)	544 (4.8)	161 (1.3)
Economics and business	20,776 (1.4)	93 (0.6)	240 (2.1)	0
Enabling and strategic technologies	101,820 (6.8)	1191 (8.0)	1222 (10.8)	878 (7.3)
Engineering	56,200 (3.8)	694 (4.6)	1374 (12.2)	141 (1.2)
General arts, humanities, and social sciences	2 (<0.1)	0	0	0
General science and technology	77 (<0.1)	0	0	0
Historical studies	1550 (0.1)	12 (0.1)	27 (0.2)	0
Information and communication technologies	57,163 (3.8)	159 (1.1)	1240 (11.0)	191 (1.6)
Mathematics and statistics	9605 (0.6)	72 (0.5)	1065 (9.4)	5 (<0.1)
Philosophy and theology	433 (<0.1)	0	18 (0.2)	0
Physics and astronomy	174,028 (11.6)	1462 (9.8)	NA	NA
Psychology and cognitive sciences	18,264 (1.2)	126 (0.8)	224 (2.0)	30 (0.3)
Public health and health services	19,928 (1.3)	138 (0.9)	108 (1.0)	81 (0.7)
Social sciences	12,372 (0.8)	69 (0.5)	189 (1.7)	4 (<0.1)
Visual and performing arts	20 (<0.1)	0	2 (<0.1)	0

Abbreviation: NA, not applicable.

[a]Excluding physics and astronomy due to the influence of nuclear and particle physics.

2. Waltman L, van Eck NJ, van Leeuwen TN, Visser MS, van Raan AFJ. Towards a new crown indicator: some theoretical considerations. *J Informetrics*. 2011;5(1):37-47. doi:10.1016/j.joi.2010.08.001

[1]LIX, École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France; [2]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, US, jioannid@stanford.edu; [3]Department of Informatics and Telecommunications, University of Peloponnese, Tripoli, Greece; [4]Department of Computer Science, University of Luxembourg, Esch-sur-Alzette, Luxembourg; [5]Mohamed bin Zayed University of Artificial Intelligence, Masdar City, Abu Dhabi, United Arab Emirates.

Conflict of Interest Disclosures John P. A. Ioannidis 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.

Additional Information Iakovos Evdaimon is a co–corresponding author (ievdaimon7@gmail.com).

Scale and Resilience in Organizations Enabling Systematic Scientific Fraud

Reese A. K. Richardson,[1,2] Spencer S. Hong,[1,3] Jennifer A. Byrne,[4,5] Thomas Stoeger,[6,7,8] Luís A. Nunes Amaral[1,2,9,10,11]

Objective Some suggest that the ease of communication provided by the internet and open access publishing has created the conditions for the emergence of entities (including paper mills, brokers, and predatory publishers) that facilitate systematic scientific fraud. However, little is understood about the organization of these entities as well as how they react to and evade science integrity measures. Here, we sought to demonstrate that large networks that produce scientific fraud at scale can be identified by the footprints they have left in the published scientific literature. These footprints can be analyzed to glean insights about their organizational structure and operations.

Design Our work consisted of observational case studies making use of article metadata from OpenAlex and the complete corpora of *PLOS One* and Hindawi articles; all PubPeer comments made before January 2024; historical indexing data from MEDLINE, Scopus, and Web of Science; and archived webpages obtained with the Internet Archive's Wayback Machine. Our case studies addressed networks of compromised editors at large mega-journals, analysis of article provenance through observations of interarticle image duplication, a longitudinal study of the operations of a large broker organization, a comparison of revision and retraction rates in closely related biomedical subfields, and an overall assessment of growth rates of systematic scientific fraud. This analysis was performed from June to September 2024.

Results We identified 45 editors at *PLOS One*, 53 editors at 10 Hindawi journals, and 205 Institute of Electrical and Electronics Engineers (IEEE) conferences that handled articles that were eventually retracted, articles with PubPeer comments, or articles featuring tortured phrases far more often than expected by chance (1-sided Poisson binomial test, Benjamini-Hochberg false discovery rate, <0.05). We characterized a network of image duplication spanning 2213 articles. We catalogued the involvement of a broker organization with an evolving portfolio spanning 188 journals over 7 years. We found similar rates of revision in subfields of RNA biology but retraction rates orders of magnitude apart. We found that while publication rates of scientific articles doubled approximately every 15 years and the publication rates of eventually retracted articles every 3 years, articles of likely paper mill provenance were published at a rate that doubled less than every 2 years.

Conclusions Here, we demonstrated through case studies that (1) individuals have colluded to publish problematic papers in a number of journals; (2) brokers can ensure publication in infiltrated journals at scale; and (3) within a field of science, not all subfields are equally targeted for scientific fraud. Our results revealed some of the strategies that enable entities promoting scientific fraud to evade interventions. Our final analysis suggested that this ability to evade interventions is enabling the number of fraudulent publications to grow at a rate far outpacing that of legitimate science.

[1]Department of Engineering Sciences & Applied Mathematics, Northwestern University, Evanston, IL, US, richardsonr43@gmail.com; [2]Department of Molecular Biosciences, Northwestern University, Evanston, IL, US; [3]Department of Chemical & Biological Engineering, Northwestern University, Evanston, IL, US; [4]School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, NSW, Australia; [5]NSW Health Statewide Biobank, NSW Health Pathology, Camperdown, NSW, Australia; [6]The Potocsnak Longevity Institute, Northwestern University, Chicago, IL, US; [7]Simpson Querrey Lung Institute for Translational Science, Northwestern University, Chicago, IL, US; [8]Division of Pulmonary and Critical Care, Northwestern University Feinberg School of Medicine, Chicago, IL, US; [9]Department of Physics and Astronomy, Northwestern University, Evanston, IL, US; [10]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, US; [11]NSF-Simons National Institute for Theory and Mathematics in Biology, Chicago, IL, US.

Conflict of Interest Disclosures Luís A. Nunes Amaral is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Funding/Support Reese A. K. Richardson was supported in part by a National Institutes of Health Training Grant (T32GM008449) through Northwestern University's Biotechnology Training Program and receives funding from the Dr John N. Nicholson Fellowship from Northwestern University and Moderna. Spencer S. Hong receives support from the Ryan Fellowship and the International Institute for Nanotechnology at Northwestern University. Jennifer A. Byrne receives funding from a National Health and Medical Research Council of Australia Ideas grant (APP1184263) and Rewarding Research Success funding from the Faculty of Medicine and Health at The University of Sydney. Thomas Stoeger receives funding from the National Institute on Aging, Integrative Multi-Scale Systems Analysis of Gene-Expression-Driven Aging Morbidity (R00AG068544), National Institute of Allergy and Infectious Diseases (AI135964), and Successful Clinical Response In Pneumonia Therapy (SCRIPT) Systems Biology Center. Luís A. Nunes Amaral and Thomas Stoeger receive funding from SCISIPBIO: a data-science approach to evaluating the likelihood of fraud and error in published studies (1956338).

Role of the Funder/Sponsor The funders had no role in the design or execution of this study.

Patterns of Paper Mill Papers and Retraction Challenges

Anna Abalkina,[1] Svetlana Kleiner[2]

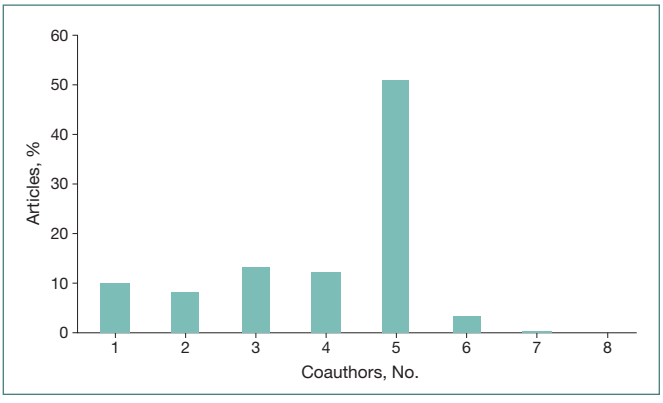
Objective Paper mills pose a significant challenge to the scientific community, yet knowledge of paper mills remains fragmented. An estimated 400,000 studies originate from paper mills,[1] although only 56,000 have been retracted or corrected. This study investigated the patterns of paper-mill articles and the challenges of their retraction, using a sample from Tanu.pro,[2] one of Europe's largest paper mills.

Design Tanu.pro creates unique emails to submit papers, often leading to discrepancies between the country domains and authors' affiliations. Employing a snowball method, the study identified more than 60 suspicious email domains, such as tanu.pro (which gave the paper mill its title in the literature). To retrieve Tanu.pro articles, we conducted a full-text search in Scopus on these domains on December 14, 2024. Combining both outsider and insider perspectives, we analyzed the retraction challenges, focusing on Tanu.pro papers published in Springer journals. A sample of 77 Springer articles was used to examine scholars' positions and compare the results with a control group of 77 randomly selected articles without paper-mill hallmarks but with the matching indicators (year, journal, number of coauthors). The ranks and positions of all authors in both groups were examined, where possible, at the time of submission.

Results The study detected 1517 papers published between 2017 and 2025 in 380 journals and coauthored by more than 4500 scholars from 46 countries and over 460 universities. While a problematic email was not definitive proof of malpractice, there was evidence that some Tanu.pro papers were associated with an author number anomaly (**Figure 25-1113**), poor empirical analysis, irrelevant citations, data fabrication, errors, fake reviewer suggestions, compromised special issues, and translated plagiarism. The Springer articles were initially identified by authors' emails; an additional selection criterion was the presence of suggested reviewers. Tanu.pro papers had a higher proportion of scholars occupying top university positions, such as professors or department heads, than in the control group (44% vs 30%). A problematic email as the only evidence posed challenges for action by journals when other issues are absent: 20 of the 77 Tanu.pro articles (26%) could not be retracted under current Committee on Publication Ethics (COPE) guidelines. Editors also felt more reluctant to retract when the only red flags were emails and unused suggested reviewers.

Conclusions The study has shown that specific paper mill patterns are useful when investigating problematic articles. However, some red flags are insufficient grounds for paper

Figure 25-1113. Mean Number of Coauthors



retraction due to limited clarity of COPE guidelines that place some articles in the gray zone if they do not feature strong grounds for retraction. More retractable categories are necessary so that COPE members can mark dubious research more easily, eg, undisclosed third-party submission. Although our recommendations are limited by the analysis of one paper mill, COPE guidelines need regular updating considering patterns of other paper mills.

References

1. Van Noorden R. How big is science's fake-paper problem? *Nature*. 2023;623(7987):466-467. doi:10.1038/d41586-023-03464-x

2. Bishop D, Abalkina A. Paper mills: a novel form of publishing malpractice affecting psychology. *Meta-Psy*. March 17, 2024. https://open.lnu.se/index.php/metapsychology/article/view/3422

[1]Freie Universität Berlin, Berlin, Germany, abalkina@gmail.com; [2]Springer Nature, Dordrecht, the Netherlands.

Conflict of Interest Disclosures None reported.

Sustainable Approaches to Upholding High Integrity Standards in the Face of Large-Scale Threats: Insights From *PLOS One*

Renee V. Hoch,[1] Emily J. Chenette[1]

Objective *PLOS One* has experienced a sharp rise in paper mills and poor-quality submissions since 2021 and has taken an agile, multifaceted preventive approach to address these issues at scale. This quality improvement study evaluated trends in large-scale integrity issues (ie, megacases) at *PLOS One* between 2021 and 2024 and the impacts of preventive initiatives on the journal's ability to address megacase concerns prior to publication.

Design We analyzed temporal and geographic trends, issue types, and research areas for megacases detected at *PLOS One* between 2021 and 2024. We then evaluated changes to desk reject rates following specific interventions as well as changes to the pre- vs postpublication impacts of megacases over time.

Results Between 2021 and 2024, *PLOS One* detected 50 megacases that cumulatively involved more than 7000 submissions and publications, with concerns including paper mills (42%; 21 of 50), authorship (22%; 11 of 50), peer review (52%; 26 of 50), and policy compliance (12%; 6 of 50). Megacases spanned multiple research areas, with the highest representation in the physical sciences and engineering (42%; 21 of 50) and medicine and public health (18%; 9 of 50). Most megacases involved submissions from geographically clustered author groups; megacases were received from 19 countries, with the highest representation from Pakistan (20%; 10 of 50), China (18%; 9 of 50), and India (10%; 5 of 50). The number of megacases detected by PLOS increased from 2 (214 articles) in 2021 to 25 (4310 articles) in 2024, and during this time the journal's desk reject rate more than doubled, from 13% (5188 of 40,548) in 2021 to 27% (16,160 of 59,425) in 2024. In 2024 alone, we rejected more than 3000 submissions flagged for megacase concerns. Desk reject data in periods before and after specific preventive interventions indicated that each of these efforts increased the journal's ability to detect and/or reject concerning content (**Table 25-0967**). We quantified pre- vs postpublication megacase impacts using a preventive efficacy metric, defined as the number of submissions flagged for megacase concerns (ie, prepublication impact) divided by the number of affected publications (ie, postpublication impact). *PLOS One*'s preventive efficacy increased from 0.70 in 2022 to 9.09 in 2024; the cumulative number of affected articles increased substantially during this time, but articles of concern were increasingly rejected prepublication, and the postpublication impact actually decreased (530 affected publications in 2022 vs 427 affected publications in 2024).

Conclusions Large-scale integrity issues have risen considerably in recent years, affecting a broad array of subject areas and geographies. Through varied approaches, *PLOS One* has improved prepublication detection and rejection of content linked to megacases. This reduces the volume of unreliable publications, the resources required to address these integrity issues, and the impacts of megacases on peer review. Such scalable preventive approaches are crucial to the long-term sustainability of integrity work in an era of increasing paper mill activity.

[1]PLOS (Public Library of Science), Cambridge, UK, and San Francisco, CA, US, rhoch@plos.org.

Conflict of Interest Disclosures Renee V. Hoch is an employee of PLOS; the head of the PLOS Publication Ethics team; a member of the Committee on Publication Ethics (COPE) Council, a United2Act Working Group, and an STM Task & Finish Working Group that developed guidance pertaining to large-scale integrity cases; a former member of the COPE Paper Mills Working Group; and a coauthor of COPE guidance on addressing large-scale paper mill issues. Emily J. Chenette is an employee of PLOS and is the editor in chief of *PLOS One* and head of PLOS editorial board services.

Acknowledgment We thank the *PLOS One* Editorial and Publication Ethics teams for their work identifying and addressing large-scale cases, the PLOS editorial board management team for their help investigating and offboarding editorial board members of concern, and the PLOS Digital team for their help developing data analytics and technology tools needed for this work. Thanks also to Suzanne Farley, Rebecca Kirk, Zena Nyakoojo, and Dena Emmerson for their input on this abstract.

Additional Information PLOS data and resources were used for this study, and others within PLOS Editorial provided input and contributed to the work underlying the datasets. PLOS was not otherwise involved in the reported study.

Table 25-0967. Interventions[a] and Outcomes at *PLOS One*

Intervention	Implementation date	Desk reject rate before intervention	Desk reject rate after intervention	No. of rejections
Manual title scanning (all submissions) to identify megacase manuscripts	Mid-2022	2022 Q1-Q2 desk reject rate: 12% (2164/18,413)	2022 Q3-Q4 desk reject rate: 15% (2544/16,775)	Desk rejected 184 manuscripts submitted in 2022 that were flagged for megacase concerns
Improved reporting-based methods of detecting submissions associated with known megacases and a new policy on manipulation of the publication process supported rejection or retraction if there were concerns about paper mill activity or related issues	2023	In 2022, 3.9% (184/4708) of desk rejected submissions were flagged for megacase concerns, and 50% (184/372) of submissions flagged for megacase concerns were desk rejected	In 2023, 8.7% (771/8854) of desk rejected submissions were flagged for megacase concerns, and 58% (771/1321) of submissions flagged for megacase concerns were desk rejected	Desk rejected 771 manuscripts submitted in 2023 that were flagged for megacase concerns
Policy update to require ethics documentation for human research	March 2023	18% (681/3853) Of human research submissions received Jan-Feb 2023	29% (1599/5422) Of human research submissions received Mar-May 2023	NA
STM Integrity Hub duplicate submission tool	December 2023	NA	NA	873 Submissions
Enhanced study design and reporting checks for MRSs	March 2024	29% (153/536) Of MRSs received Mar 2023-Feb 2024	85% (1690/1996) Of MRSs received Mar 2024-Feb 2025	NA
Enhanced study design and reporting checks for SRMAs	December 2024	18% (686/3742) Of SRMAs received Jan 2024-Dec 2024	88% (1380/1562) Of SRMAs received Dec 2024-May 2025	NA
Pilot audits to detect submissions associated with concerning contributor accounts	Mid-2024	NA	NA	666 Submissions (2 megacases)[b]
Total annual desk reject rate	2021: 13% (5188/40,548)	2022: 13% (4708/35,188)	2023: 20% (8854/43,702)	2024: 27% (16,160/59,425)

Abbreviations: MRS, Mendelian randomization study; NA, not applicable; SRMA, systematic review and meta-analysis.

[a]To safeguard PLOS' processes, we are not providing details of screening and audit approaches because we must ensure that information that could be used to circumvent our processes does not get into the public or published domain. We would be happy to provide more information in confidential discussions with interested journals or publishers.

[b]This represents only a subset of pilot audits conducted in 2024.

Thursday, September 4

Bias, Study Outcomes, and Reporting Concerns

Immortal Time Bias Prevalence and Influence on Estimates in Systematic Reviews and Meta-Analyses

Jae Il Shin,[1,2,3] Min Seo Kim,[4,5,6,7] Dong Keon Yon,[8] Seung Won Lee,[9] Masoud Rahmati,[10] Marco Solmi,[11,12,13,14,15] Andre F. Carvalho,[16] Ai Koyanagi,[17,18,19] Lee Smith,[20] John P. A. Ioannidis[21,22,23,24,25]

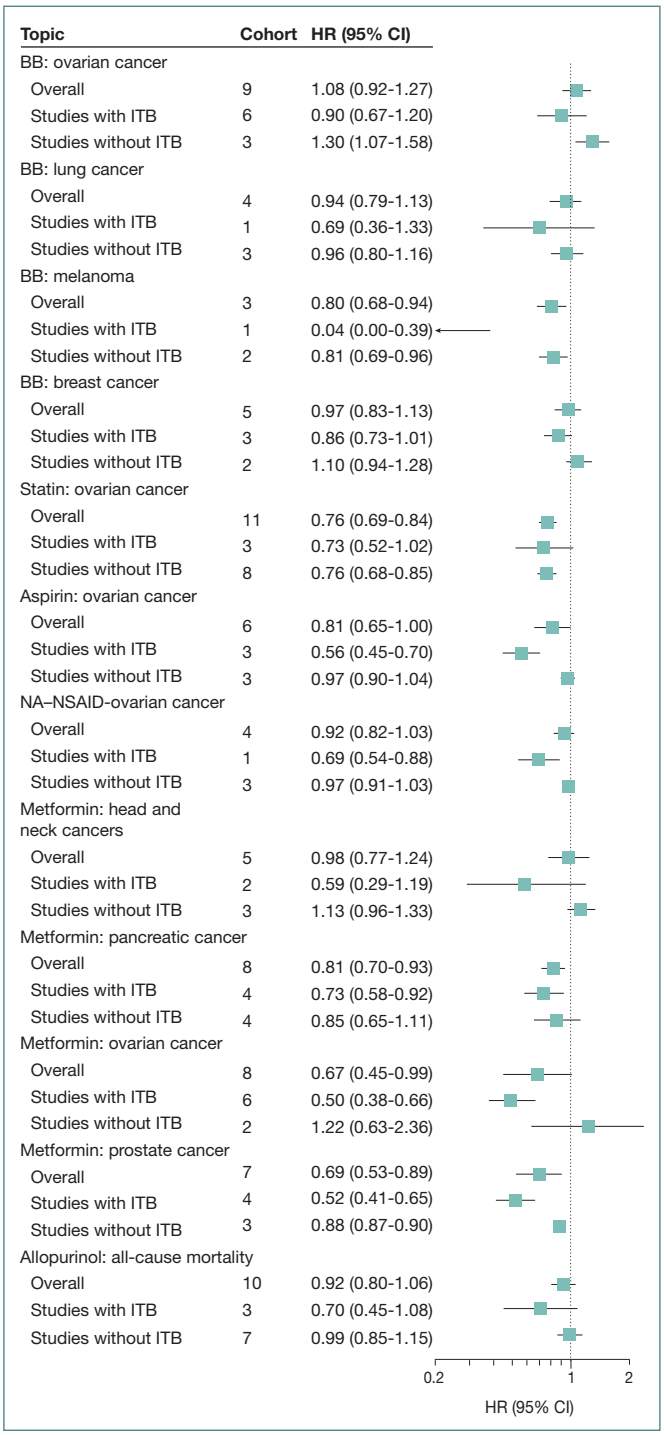
Objective Immortal time bias (ITB) is the error in estimating the association between the exposure and the outcome that results from misclassification or exclusion of time intervals. The aim of our study was to estimate the prevalence of ITB in systematic reviews and meta-analyses and to assess the degree to which it contributes to effect size estimates and evidence reversal.

Design We performed a systematic review of systematic reviews with meta-analyses (SRMA) that only underwent detailed analysis based on ITB. We only included SRMA of observational studies including cohort and case-control studies and excluded those without meta-analysis or SRMA of randomized control trials. We searched PubMed/MEDLINE, Embase, and Cochrane Database of Systematic Reviews from database inception to July 31, 2024. Two authors independently extracted data and evaluated the methodological quality of the systematic reviews. Information on ITB judgment and effect sizes with 95% CIs for individual studies in forest plots were extracted to run reanalysis using generic inverse variance fixed- and random-effects methods. After extracting data, we conducted subgroup analyses by the presence of ITB for all available topics and assessed the influence of ITB on the heterogeneity (I^2), changes of evidence, statistical significance of the finding, and altering effect size in favor of intervention/exposure.

Results Among the 12 systematic reviews with relevant data, there were 25 eligible topics (only 21 could be divided by ITB, as 4 included only studies without ITB). The median (IQR) number of studies included for a topic was 6 (4-10). Among 182 studies among 25 topics, 44.0% (80 studies) were affected by ITB. Among the 21 topics where both studies with ITB and studies without ITB were available, 57.1% (12/21) demonstrated discordant results between ITB subgroups (**Figure 25-0976**). Evidence reversal occurred in 23.8% (5/21), where overall summary results changed from statistically significant to non–statistically significant or vice versa after excluding studies with ITB. The ratio of effect size (effect sizes pooled from studies with ITB relative to those pooled from studies without ITB) was 0.71 (95% CI, 0.66-0.78), suggesting that the effect sizes from studies with ITB were exaggerated by an average of 29% in favor of the intervention/exposure. Excluding studies involving ITB reduced the heterogeneity (I^2) of overall pooled results by 21.4% on average.

Conclusions We quantitatively captured how far ITB has influenced our knowledge and clinical practices. Given the projected high prevalence and nontrivial influence of ITB, ITB should be considered in studies with survival analyses, and

Figure 25-0976. Overall and Subgroup Analysis by Immortal Time Bias (ITB) for the Associations Between Drugs and Overall Survival



BB indicates β-blocker; HR, hazard ratio; NA-NSAID, nonaspirin nonsteroidal anti-inflammatory drug.

improving reporting standards by researchers as well as collective surveillance from readers, reviewers, and editors is warranted. Future studies should address how ITS may also interact with other trial characteristics and biases in affecting treatment effect estimates.

[1]Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea, shinji@yuhs.ac; ; [2]Severance Underwood Meta-Research Center, Institute of Convergence Science, Yonsei University, Seoul, Republic of Korea; [3]Affiliate in Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, US; [4]Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea; [5]Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute, Cambridge, MA, US; [6]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, US; [7]Department of Medicine, Harvard Medical School, Boston, MA, US; [8]Center for Digital Health, Medical Science Research Institute, Kyung Hee University College of Medicine, Seoul, Republic of Korea; [9]Department of Data Science, Sejong University College of Software Convergence, Seoul, Republic of Korea; [10]Department of Physical Education and Sport Sciences, Faculty of Literature and Human Sciences, Lorestan University, Khoramabad, Iran; [11]Department of Psychiatry, University of Ottawa, Ottawa, Ontario, Canada; [12]Department of Mental Health, The Ottawa Hospital, Ottawa, Ontario, Canada; [13]Ottawa Hospital Research Institute (OHRI) Clinical Epidemiology Program University of Ottawa, Ottawa, Ontario, Canada; [14]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada; [15]Department of Child and Adolescent Psychiatry, Charité Universitätsmedizin, Berlin, Germany; [16]Innovation in Mental and Physical Health and Clinical Treatment, Strategic Research Centre, School of Medicine, Barwon Health, Deakin University, Geelong, VIC, Australia; [17]Research and Development Unit, Parc Sanitari Sant Joan de Déu, Universitat de Barcelona, Fundació Sant Joan de Déu, CIBERSAM, Barcelona, Spain; [18]ICREA, Pg. Lluis Companys 23, Barcelona, Spain; [19]Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM, Madrid, Spain; [20]The Cambridge Centre for Sport and Exercise Sciences, Anglia Ruskin University, Cambridge, United Kingdom; [21]Department of Medicine, Stanford University, Stanford, CA, US; [22]Department of Epidemiology and Population Health, Stanford University, Stanford, CA, US; [23]Department of Biomedical Data Science, Stanford University, Stanford, CA, US; [24]Department of Statistics, Stanford University, Stanford, CA, US; [25]Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, US.Severance Underwood Meta-Research Center, Institute of Convergence Science, Yonsei University, Seoul, Republic of Korea; [3]Affiliate in Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, US; [4]Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea; [5]Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute, Cambridge, MA, US; [6]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, US; [7]Department of Medicine, Harvard Medical School, Boston, MA, US; [8]Center for Digital Health, Medical Science Research Institute, Kyung Hee University College of Medicine, Seoul, Republic of Korea; [9]Department of Data Science, Sejong University College of Software Convergence, Seoul, Republic of Korea; [10]Department of Physical Education and Sport Sciences, Faculty of Literature and Human Sciences, Lorestan University, Khoramabad, Iran; [11]Department of Psychiatry, University of Ottawa, Ottawa, Ontario, Canada; [12]Department of Mental Health, The Ottawa Hospital, Ottawa, Ontario, Canada; [13]Ottawa Hospital Research Institute (OHRI) Clinical Epidemiology Program University of Ottawa, Ottawa, Ontario, Canada; [14]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada; [15]Department of Child and Adolescent Psychiatry, Charité Universitätsmedizin, Berlin, Germany; [16]Innovation in Mental and Physical Health and Clinical Treatment, Strategic Research Centre, School of Medicine, Barwon Health, Deakin University, Geelong, VIC, Australia; [17]Research and Development Unit, Parc Sanitari Sant Joan de Déu, Universitat de Barcelona, Fundació Sant Joan de Déu, CIBERSAM, Barcelona, Spain; [18]ICREA, Pg. Lluis Companys 23, Barcelona, Spain; [19]Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM, Madrid, Spain; [20]The Cambridge

Centre for Sport and Exercise Sciences, Anglia Ruskin University, Cambridge, UK; [21]Department of Medicine, Stanford University, Stanford, CA, US; [22]Department of Epidemiology and Population Health, Stanford University, Stanford, CA, US; [23]Department of Biomedical Data Science, Stanford University, Stanford, CA, US; [24]Department of Statistics, Stanford University, Stanford, CA, US; [25]Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, US.

Conflict of Interest Disclosures John P. A. Ioannidis is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Funding/Support The work of John P. A. Ioannidis is supported by an unrestricted gift from Sue and Bob O'Donnell to Stanford University. Jae Il Shin is supported by the Yonsei Faculty Fellowship, funded by Lee Youn Jae.

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

Additional Information John P. A. Ioannidis is a co–corresponding author (jioannid@stanford.edu).

Effect Estimates for the Same Outcomes Designated as Primary vs Secondary in Randomized Clinical Trials: A Meta-Research Study

Yuanxi Jia,[1] Yiwen Jiang,[2] Karen A. Robinson,[3] Jinling Tang[4]

Objective Researchers conducting randomized clinical trials (RCTs) might focus more on primary outcomes than secondary outcomes. Therefore, when the design and conduct of an RCT were manipulated for positive results, the effect size of the primary outcome may be overestimated more than that of the secondary outcomes. Among RCTs with similar design, the discrepant effect estimates of the same outcome between those designating it as a primary outcome (PO-RCTs) and others designating it as a secondary outcome (SO-RCTs) may indicate the impact of bias. This study aimed to compare the effect estimates of the same outcomes between PO-RCTs and SO-RCTs. PO-RCTs were hypothesized to produce higher effect estimates than SO-RCTs.

Design This meta-research study using a matched cohort design included 153 meta-analyses assessing the efficacy/effectiveness of health interventions published in the Cochrane Database of Systematic Reviews between 2021 and 2023. A total of 1073 RCTs with parallel design, individual randomization, superiority tests, recruitment in or after 2006, and negative control (placebo, sham, no intervention, waiting list, minimal intervention, attention control, or standard care) were included. Within each meta-analysis, RCTs designating the meta-analyzed outcome as primary outcome were classified as PO-RCTs (553 RCTs), while those designating the meta-analyzed outcome as secondary outcome were classified as SO-RCTs (520 RCTs). RCTs that did not define a primary outcome were excluded. The PO-RCTs were compared with the SO-RCTs using 2-stage random-effect meta-analyses: the effect estimates from RCTs were transformed into odds ratios (ORs); within each meta-analysis, the ORs from PO-RCTs and SO-RCTs were compared as a ratio of ORs (ROR); and the RORs across meta-analyses were combined. The primary analyses were conducted among all RCTs and prospectively registered RCTs. The risk of bias for blinding was assessed using the Cochrane Risk of Bias Tool.

Results Among 1073 RCTs from 153 meta-analyses, PO-RCTs produced ORs 1.27 (95% CI, 1.14-1.42; I^2 = 43.6%) times higher than SO-RCTs. Among 372 prospectively registered RCTs from 122 meta-analyses, PO-RCTs produced ORs 1.21 (95% CI, 1.03-1.41; I^2 = 22.3%) times higher than SO-RCTs. When restricted to RCTs with high/unclear risk in performance bias, detection bias, or both types of bias, PO-RCTs produced ORs 1.48 (95% CI, 1.26-1.75), 1.44 (95% CI, 1.21-1.71), and 1.60 (95% CI, 1.30-1.96) times higher than SO-RCTs, respectively. When restricted to RCTs with low risk in performance bias, detection bias, or both types of bias, PO-RCTs produced ORs 1.08 (95% CI, 0.90-1.28), 1.04 (95% CI, 0.87-1.24), and 1.06 (95% CI, 0.90-1.26) times higher than SO-RCTs, respectively.

Conclusions The effect estimate for an outcome could be 27% larger when designated as a primary outcome than a secondary outcome in RCTs, which may be partially attributed to bias. Researchers conducting systematic reviews may need to note whether an outcome was designated as primary or secondary in RCTs, perform sensitivity analyses, and interpret results considering the impact of potential bias.

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, yx.jia@nus.edu.sg; [2]Shenzhen Institute of Advanced Technology, Shenzhen, China; [3]School of Medicine, Johns Hopkins University, Baltimore, US; [4]Shenzhen University of Advanced Technology, Shenzhen, China.

Conflict of Interest Disclosures None reported.

Funding/Support This study was supported by the Shenzhen Science and Technology Program (grant KQTD20190929172835662) from the Shenzhen Municipal Government, Guangdong Province, China, and the Outstanding Youth Innovation Fund from the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (grant E2G019).

Role of the Funder/Sponsor: The funders had no roles in the design, conduct, or reporting of this study.

Detection and Monitoring of Outcome Reporting Changes Using a Large Language Model: Application to FDA-Regulated Drug Trials

Ian Bulovic,[1] Susmitha Wunnava,[1,2] Wonjin Yoon,[1,3] Adam G. Dunn,[1,4] Florence T. Bourgeois,[1,2,3] Timothy Miller[1,2,3]

Objective Selective inclusion and reporting of clinical trial outcomes may result in biases in the medical evidence on benefits and harms of treatments.[1] ClinicalTrials.gov provides the opportunity to identify changes between prespecified outcomes and those eventually reported in publications, but linkage and comparison of these outcomes requires extensive manual curation, precluding studies of more than a few hundred trials.[2,3] ClinicalTrials.gov archives all versions of registration records as changes are made by investigators,

providing the basis to monitor trial outcome changes over time if the information could be efficiently processed. Our objective was to use a large language model (LLM) to estimate changes to primary outcomes at scale.

Design OpenAI's Gpt-4o-2024-08-06 model was prompted (without training) to identify trials with meaningful changes in specified primary outcomes between trial registrations at study start and in the final record (prompts and scripts used for processing data available at https://github.com/Machine-Learning-for-Medical-Language/outcome-switching. Meaningful changes were defined as addition, removal, or substantial modification to primary outcomes. Substantial modifications included changes to outcome measures, measurement procedures, or measurement time frame. The model performance was evaluated using a set of 100 manually labeled and adjudicated records, showing F1 scores of 0.88 for outcome addition, 0.81 for outcome removal, and 0.67 for outcome modification (F1 scores range from 0 to 1 and balance sensitivity and positive predictive value). The cohort consisted of interventional trials studying US Food and Drug Administration (FDA)–regulated drugs that were prospectively registered and started January 1, 2008, to December 31, 2022 (allowing ≥2 years' follow-up). Variables of interest (start year, therapeutic area, industry funding, randomization, and enrollment size) were selected a priori and extracted from registration records using structured data available for download from ClinicalTrials.gov. Univariate analyses and multivariate logistic regression were performed to assess for associations between outcome changes and variables of interest.

Results Among 27,227 trials studying FDA-regulated drugs, 11,078 (40.7%) had changes to primary outcomes after trial start. This included 3850 trials (14.1%) with addition of an outcome, 3123 (11.5%) with removal of an outcome, and 9105 (33.4%) with substantial outcome modifications. Trials with industry funding were significantly more likely to have an outcome change (47.7% vs 32.9%; $P < .001$) (**Table 25-1101**). This difference was observed for each of the outcome change types, including addition (17.4% vs 10.4%; $P < .001$), removal (12.9% vs 9.9%; $P < .001$), and modification (39.7% vs 26.4%; $P < .001$). Multivariate analysis demonstrated reduction in outcome changes over time ($P < .001$) and a positive association with industry funding ($P < .001$).

Conclusions LLMs can be used to monitor for changes in clinical trial outcomes efficiently and at scale using information available in ClinicalTrials.gov. Outcome changes in trials of FDA-regulated drugs are common and associated with industry funding but have decreased over time. Future work should explore use of LLMs to identify outcome changes between trial registrations and published articles.

Table 25-1101. Characteristics of 27,227 US Food and Drug Administration–Regulated Drug Trials Registered in ClinicalTrials.gov

	Total trials, No. (%)	Trials with primary outcome changes, No. (%)	*P* value[a]
Start year			**<.001**
2008-2010	330 (1.2)	272 (82.4)	
2011-2013	882 (3.2)	648 (73.5)	
2014-2016	2261 (8.3)	1519 (67.2)	
2017-2019	11,746 (43.1)	5141 (43.8)	
2020-2022	12,008 (44.1)	3498 (29.1)	
Therapeutic area[b]			**<.001**
Neoplasms	10,168 (37.3)	4529 (44.5)	
Immune system diseases	3962 (14.6)	1796 (45.3)	
Hematologic and lymphatic diseases	3495 (12.8)	1605 (45.9)	
Respiratory tract diseases	3151 (11.6)	1435 (45.5)	
Skin and connective tissue diseases	3010 (11.1)	1281 (42.6)	
Urogenital diseases	2968 (10.9)	1309 (44.1)	
Infections	2950 (10.8)	1308 (44.3)	
Nervous system diseases	2823 (10.4)	1184 (41.9)	
Other[c]	8159 (30.0)	2801 (34.3)	
Any industry funding			**<.001**
Yes	14,404 (52.9)	6865 (47.7)	
No	12,823 (47.1)	4213 (32.9)	
Randomized study design			**.34**
Yes	14,431 (53.0)	5811 (40.3)	
No	12,796 (47.0)	5267 (41.2)	
Trial enrollment, No. of participants			**<.001**
0-19	6397 (23.5)	2128 (33.3)	
20-49	7065 (25.9)	2545 (36.0)	
50-99	4735 (17.4)	1982 (41.9)	
100-499	6794 (25.0)	3209 (47.2)	
≥500	2236 (8.2)	1214 (54.3)	

[a]For univariate analyses using χ^2 tests to assess for an association between the variable of interest and the presence of outcome changes.

[b]Trials could be assigned to more than 1 therapeutic area based on condition MeSH (Medical Subject Headings) terms available in ClinicalTrials.gov.

[c]Included therapeutic areas with less than 10% of trials.

2. Wang A, Menon R, Li T, et al. Has the degree of outcome reporting bias in surgical randomized trials changed? a meta-regression analysis. *ANZ J Surg*. 2023;93:76-82. doi:10.1111/ans.182733

3. Hinkel J, Heneghan C, Bankhead C. Selective outcome reporting in cancer studies: a scoping review. *medRxiv*. 2024:07.02.24309826. doi:10.1101/2024.07.02.24309826

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, US, florence_bourgeois@hms.harvard.edu; [2]Harvard-MIT Center for Regulatory Science, Harvard Medical School, Boston, MA, US; [3]Department of Pediatrics, Harvard Medical School, Boston, MA, US; [4]Biomedical Informatics and

References

1. Turner EH, Cipriani A, Furukawa TA, Salanti G, de Vries YA. Selective publication of antidepressant trials and its influence on apparent efficacy: updated comparisons and meta-analyses of newer versus older trials. *PLoS Med*. 2022;19(1):e1003886. doi:10.1371/journal.pmed.10038862

Digital Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia.

Conflict of Interest Disclosures None reported.

Funding/Support National Library of Medicine, National Institutes of Health (R01LM012976, R01LM012973).

Role of the Funder/Sponsor The funder played no role in the design or reporting of the study.

Additional Information Florence T. Bourgeois and Timothy Miller are co–senior authors.

Data Repurpose in AI Studies and Scientific Outcomes

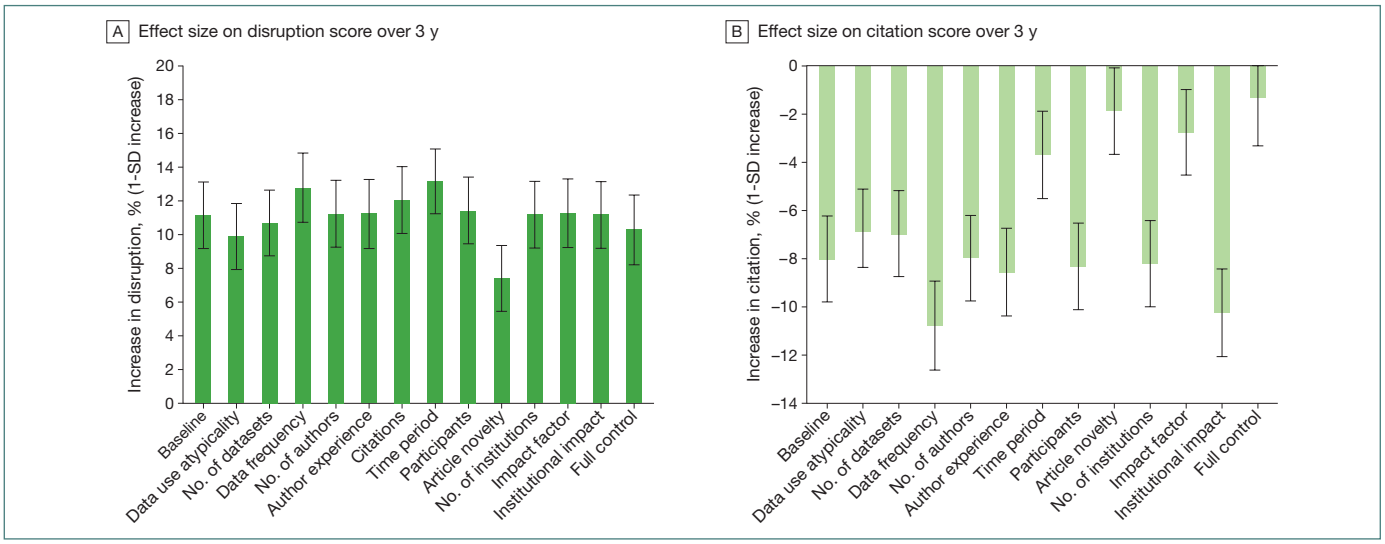
Yulin Yu,[1] Yong-Yeol Ahn,[2] Daniel Romero[1,2,3,4]

Objective Datasets are critical in artificial intelligence (AI) research, serving as the foundation for training, evaluating, and developing models. As interest in accelerating AI-driven discoveries grows, it becomes important to examine how strategically repurposing datasets, using them for research topics different from their original context, relates to scientific outcomes. Previous research suggests that combining existing knowledge in novel ways, or recombinant novelty, is associated with greater scientific influence.[1,2] While recent studies have shown that combining datasets in unexpected ways correlates with impact, they have not examined how datasets interact with the topics they are used to study.[3] To address this, we draw on transformational creativity theory, which modifies the accepted conceptual space by altering or removing existing dimensions, such as drug repurposing. Similarly, AI datasets can be repurposed for new research topics. This study addressed 2 questions: (1) are published articles that repurpose datasets associated with higher scientific disruption or recognition? and (2) do repurposed uses of datasets influence how future articles apply the data, and is this repurpose adoption associated with greater scientific outcomes? We conceptualized scientific impact along 2 dimensions: (1) the extent to which an article is recognized and immediately used by research communities, which can be measured by citations, and (2) the degree to which the research article disrupts existing research paradigms.

Design We analyzed 13,637 published AI articles collected from Papers With Code and linked to the OpenAlex database from 2015 to 2021. We quantified data repurposing by comparing the semantic content of each article to previous articles using the same datasets. For each article, the number of other articles using the same dataset varied (though we set a minimum threshold of at least 1 prior reuse). Articles were embedded using sentence-level language models. We assessed the relationship between repurposing and scientific disruption, defined as the extent to which an article with repurposed data is cited instead of its references, using ordinary least squares regression. We used negative binomial regression to analyze citation counts. Control variables included dataset novelty, article novelty, author information (team size and experience), publication year, citation/reference counts, and disciplinary fields. These variables help account for potential confounders; for example, more recent papers or datasets may naturally have more prior reuse by other studies. We measured repurpose adoption by comparing each article's similarity to prior vs future works using the same datasets. A positive score suggests that future articles adopted the new topical use.

Results Greater data repurposing was significantly associated with higher disruption scores, corresponding to a 10.2%-SD increase ($P < .001$) (**Figure 25-1139**, A). However, articles with repurposed data were slightly negatively associated with citation counts compared with articles without repurposed data, corresponding to a 1.7%-SD decrease, but this was not statistically significant ($P > .05$)

Figure 25-1139. Effect Size of the Data Repurposing Variable on the Disruption and Citation Scores



A, Effect size of the data repurposing variable on the disruption score over 3 years, based on an ordinary regression while controlling for various factors as indicated in the panel headings. The leftmost panels show the effect size of data repurposed in baseline regressions (without any control variables), whereas the rightmost panels display the effect size after collectively controlling for the indicated variables. B, Effect size of the data repurposing variable on the citation score over 3 years, based on a negative binomial regression while controlling for the various factors indicated in the panel headings. The leftmost panels show the effect size of data repurposed in baseline regressions (without any control variables), and the rightmost panels display the effect size after controlling for the indicated variables collectively. Error bars indicate 95% CIs.

(**Figure 25-1139**, B). Articles with 1-SD higher repurpose adoption were modestly more disruptive (+2%; $P < .001$) and showed a stronger positive association with future citations ($P < .001$). This study is limited by reliance on the OpenAlex database, which may have incomplete metadata.

Conclusions In this study, repurposing datasets was associated with greater disruption, though it may have initially reduced recognition. However, when novel data were adopted for reuse in future research, they were associated with both disruption and increased citations. These findings have implications for data sharing and research training aimed at promoting innovative reuse of data.

References

1. Uzzi B, Mukherjee S, Stringer M, Jones B. Atypical combinations and scientific impact. *Science*. 2013;342(6157):468-472. doi:10.1126/science.1240474

2. Leahey E, Lee J, Funk RJ. What types of novelty are most disruptive? *Am Sociol Rev*. 2023;88(3):562-597. doi:10.1177/00031224231168074

3. Yu Y, Romero DM. Does the use of unusual combinations of datasets contribute to greater scientific impact? *Proc Natl Acad Sci U S A*. 2024;121(15):e2318482121. doi:10.1073/pnas.2402802121

[1]School of Information, University of Michigan, Ann Arbor, US, yulinyu@umich.edu; [2]Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, US; [3]Center for the Study of Complex Systems, University of Michigan, Ann Arbor, US; [4]Computer Science and Engineering Division, University of Michigan, Ann Arbor, US.

Conflict of Interest Disclosures: None reported.

Prevalence of the Statement "to Our Knowledge" and Similar Paraphrases in Current and Past Biomedical Literature

Nicola Di Girolamo,[1,2] Reint Meursinge Reynders,[3,4] Ugo Di Girolamo[5]

Objective Statements like "to our knowledge, this is the first time that" are ambiguous,[1] but are often used when describing a potential novel aspect of an article, often in lieu of a literature search.[2] Although seemingly innocuous, similar paraphrases lack reproducibility and accountability. This study assessed the prevalence of the statement "to our knowledge" (TOK) in the biomedical literature and in leading medical journals in the past 75 years as well as before and after implementing an intervention to reduce such prevalence in 2 journals.

Design Full texts of articles available in PubMed Central (PMC) were searched using PMC's phrase index for paraphrases of the statement TOK. Prevalence was measured in articles available in PMC published by 6 leading medical journals (*The Lancet*, *NEJM*, *The BMJ*, *JAMA*, *Ann Int Med*, *PLOS Medicine*) and in all articles in PMC in 2 time periods: January 1950 to December 2019 and January 2020 to December 2024 by 1 author (N.D.G.). To evaluate the search strategy, all articles published in 2024 in *The Lancet* with full text available in PMC (N = 43) were manually screened for TOK and compared with results of the engine search. To evaluate how often TOK was reported, all *The Lancet* research articles published in 2024 (N = 166) were manually screened. In 2 journals, *Journal of Small Animal Practice* (*JSAP*) and *Journal of Exotic Pet Medicine* (*JEPM*), an intervention to advise authors against use of these phrases and in favor of a scoping search of the literature was implemented in 2020, and TOK prevalence was measured in 2018 and 2024.

Results Among 4488 articles available in PMC published in 6 leading medical journals in 2020 to 2024, the prevalence of TOK was 21.7% (974), highest in *PLOS Medicine* (48.4%) and lowest in *NEJM* (1.3%) (**Table 25-1191**). Prevalence among these journals was 5.7% (95% CI, 4.5%-6.9%) higher than all journals included in PMC. Compared with articles published in 1950 to 2019, prevalence of TOK in the last 5 years increased 8.2% (95% CI, 6.9%-9.6%) in leading medical journals and 3.1% (95% CI, 3.1%-3.2%) in all PMC journals. The search strategy had a sensitivity of 66.7% (95% CI, 34.9%-90.1%), a specificity of 100.0% (95% CI, 88.8%-100.0%), and an overall accuracy of 90.7% (95% CI, 77.9%-97.4%). In the manually screened articles, TOK was mentioned 5 times in 1 article (0.6%), 3 times in 6 articles (3.6%), 2 times in 17 articles (10.2%), and 1 time in 57 articles (34.3%). In *JSAP* and *JEPM*, TOK prevalence in articles decreased from 37.4% to 7.6% after the intervention.

Conclusions Paraphrases of TOK have been highly prevalent in the biomedical literature for the past 75 years

Table 25-1191. Prevalence of Paraphrases of the Statement "to Our Knowledge" (TOK) in Full-Text Articles Published by Journals Available in PubMed Central

Sample	Years	Articles with TOK statements	Total No. of articles	Prevalence
Leading medical journals	2020-2024	974	4488	21.7
Leading medical journals	1950-2019	2192	16,258	13.5
JAMA	2020-2024	201	1194	16.8
The BMJ	2020-2024	87	1006	8.6
NEJM	2020-2024	6	450	1.3
PLOS Medicine	2020-2024	512	1057	48.4
Ann Int Med	2020-2024	41	333	12.3
The Lancet	2020-2024	127	448	28.3
All PMC journals	2020-2024	612,981	3,831,629	16
All PMC journals	1950-2019	611,030	4,740,420	12.9
JSAP and *JEPM*	2018	26	75	34.7
JSAP and *JEPM*	2024	6	79	7.6

Abbreviations: *Ann Int Med, Annals of Internal Medicine; JEPM, Journal of Exotic Pet Medicine; JSAP, Journal of Small Animal Practice; NEJM, New England Journal of Medicine;* PMC, PubMed Central. For leading medical journals, not all journals were published for the entire time period from 1950 to 2019.

and continue to be prevalent, including in leading medical journals. A simple intervention can decrease the use of this statement. This study is limited to only the portion of articles available in PMC, does not include all articles published by the 6 leading medical journals, and does not differentiate between version types (eg, accepted manuscript, published article) or article types (eg, research, review, or opinion).

References

1. Vaisrub S. To the best of our knowledge, which is limited at best. *JAMA*. 1979;241(3):278.

2. Di Girolamo N, Meursinge Reynders R. On "authors' knowledge" and contrast-enhanced ultrasonography in rabbits. *Vet Radiol Ultrasound*. 2019;60(4):371.

[1]Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, US, nd374@cornell.edu; [2]*Journal of Small Animal Practice*, British Small Animal Veterinary Association, Gloucestershire, UK; [3]Department of Oral and Maxillofacial Surgery, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, the Netherlands; [4]private practice of orthodontics, Milan, Italy; [5]Compass, New York, NY, US.

Conflict of Interest Disclosures Nicola Di Girolamo was an editor in chief of 2 peer-reviewed journals at the time of submission, 1 published by Elsevier and 1 by Wiley. The other authors declare no competing interests nor conflicts of interest.

Peer Review Models

Peer Reviews of Peer Reviews: A Randomized Controlled Trial and Other Assessments

Alexander Goldberg,[1] Ivan Stelmakh,[2] Kyunghyun Cho,[3] Alice Oh,[4] Alekh Agarwal,[5] Danielle Belgrave,[6] Nihar B. Shah[1]

Objective What biases and errors arise in evaluating the quality of peer reviews? We studied this question, driven by 2 primary motivations: (1) designing incentive systems for high-quality reviewing based on review quality assessments and (2) evaluating experiments within peer review processes.

Design We conducted a large-scale study at the 2022 Conference on Neural Information Processing Systems (NeurIPS), a top-tier venue in machine learning, inviting reviewers, conference chairs, and authors to evaluate reviews. This study was approved by Carnegie Mellon University Institutional Review Board. Evaluators rated reviews (scores and text) on overall quality (7-point Likert scale) and individual criteria (constructiveness, coverage, understanding, and substantiation; 5-point Likert scale). To assess uselessly elongated review bias in randomized controlled trials, we created elongated versions of 10 paper reviews by adding noninformative content (duplicated instructions and copied paper abstracts), increasing length from 200-300 to 600-850 words. We randomly assigned 458 reviewers into 2 equally sized groups: the control group evaluated a short review, and the treatment group evaluated an artificially lengthened review. We used a Mann-Whitney *U* test with a test statistic that measures the proportion of paper review pairs in which the treatment (long review) was rated higher than the control (short review). To assess author outcome bias, we collected observational data from 3429 authors and 4311 paper reviewers and/or conference organizers evaluating 9870 reviews. We compared author ratings on "accept" vs "reject" reviews of their own papers, controlling for review quality by matching 418 review pairs with similar nonauthor ratings. A Mann-Whitney *U* test was conducted on these pairs. To assess interevaluator (dis) agreement, miscalibration, and subjectivity, we estimated the magnitude of various sources of error using the observational data described above.

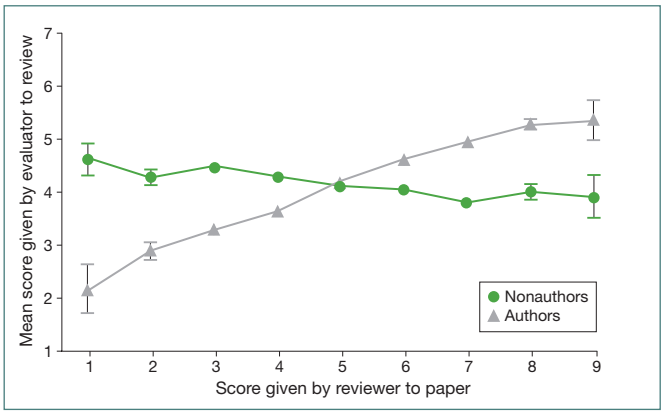
Results Evaluators exhibited a significant elongated review bias ($P < .001$). Elongated reviews received a mean score of 4.29 vs. 3.73 for original reviews. Each individual criterion also showed length bias ($P < .05$). Authors were positively biased toward reviews recommending acceptance (**Figure 24-0809**). Authors gave "reject" reviews a mean score of 3.24 compared with 4.65 on "accept" reviews. We also found a statistically significant bias in criteria scores ($P < .001$ for all criteria). The magnitude of error due to inconsistency, miscalibration, and subjectivity were similar for reviews of reviews and for reviews of papers at NeurIPS.

Conclusions Our results suggest that the various problems that exist in reviews of papers, most notably bias toward factors like length and review recommendation, also arise in reviewing of reviews. Since our experiment in 2022, our results are increasingly relevant to the evaluation of large language models in peer review settings,[1,2] where the length bias we established through our randomized controlled trial can bias results on the perceived efficacy of large language models.[3]

References

1. Thakkar N, Yuksekgonul M, Silberg J, et al. Can LLM feedback enhance review quality? a randomized study of 20K reviews at ICLR 2025. *arXiv*. Preprint posted online April 13, 2025. doi:10.48550/arXiv.2504.09737

Figure 24-0809. Review Scores Given to Papers Plotted by Mean Quality Evaluation Scores Given by Evaluators



Error bars show 95% CIs. Paper review scores range from 1 (strongest reject) to 10 (strongest accept). Authors tended to give higher scores in their evaluations toward more positive reviews on their papers.

2. D'Arcy M, Hope T, Birnbaum L, Downey D. MARG: multi-agent review generation for scientific papers. *arXiv*. Preprint posted online January 8, 2024. doi:10.48550/arXiv.2401.04259

3. Steyvers M, Tejeda H, Kumar A, et al. What large language models know and what people think they know. *Nature Machine Intelligence*. 2025;7:221-231. doi:10.1038/s42256-024-00976-7

[1]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, US, akgoldbe@andrew.cmu.edu; [2]New Economic School, Moscow, Russia; [3]Center for Data Science, New York University, New York, NY, US; [4]School of Computing, KAIST, Daejeon, South Korea; [5]Google Research, New York, NY; [6]GSK, London, UK.

Conflict of Interest Disclosures Kyunghyun Cho is affiliated with Genentech. Nihar B. Shah is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Acknowledgment We are greatly indebted to the participants of this experiment for providing evaluations of reviews, thereby helping understand the promises and challenges of evaluating review quality, and consequently also shedding light on the design of incentives and experiments in peer review.

Anonymizing Reviewers to Each Other in Peer Review Discussions: A Randomized Controlled Trial

Charvi Rastogi,[1] Xiangchen Song,[2] Zhijing Jin,[3,4] Ivan Stelmakh,[5] Hal Daumé III,[6] Kun Zhang,[2] Nihar B. Shah[2]

Objective Many peer-review processes in computer science involve reviewers submitting independent reviews followed by a discussion between reviewers of each article on a typed forum (online discussion board). A key policy question is whether reviewers should remain anonymous to each other. This study investigated 7 research questions (RQs): RQ 1. Do reviewers discuss more when anonymous to each other or not? RQ 2. Are decisions closer to senior or junior reviewers' opinions across conditions? RQ 3. Are reviewers more polite when not anonymous? RQ 4. Do self-reported reviewer experiences differ? RQ 5. Do reviewers prefer one condition? RQ 6. What factors do reviewers consider important in this policy decision? RQ 7. Have reviewers experienced dishonest behavior when their identity is revealed to other reviewers?

Design A randomized controlled trial was conducted in the Conference on Uncertainty in Artificial Intelligence (UAI) in 2022, where full articles (not abstracts) were reviewed. Reviewers and articles were randomly assigned to either a condition where reviewer identities were hidden from each other or one where they were visible. Reviewers were then matched to articles within each condition using a semiautomated procedure.[1] An anonymous survey of reviewers was also administered. The following measurements were made: RQ 1. Average posts per reviewer-article pairs were compared; test statistic: difference across conditions. RQ 2. Test statistic: difference in the fraction of articles where the reviewer closest to the final decision was senior. RQ 3. Politeness was scored 1 to 5 using a locally deployed large language model[2] with few-shot prompting; scores were averaged across iterations and paraphrased prompts; test statistic: normalized Mann-Whitney U test. RQ 4. Reviewers rated 5 aspects of their experience on a 5-point Likert scale; differences across conditions were tested using a normalized Mann-Whitney U test. RQ 5. Reviewers rated overall preference on a 5-point scale mapped from −2 to 2; test statistic: Cohen *d*. RQ 6. Reviewers rated the importance of 6 factors in deciding on anonymity policy, each from 1 (least important) to 6. RQ 7. Reviewers reported experience of any dishonest behavior due to reviewer identities being visible to other reviewers, with checkboxes "Yes, in UAI 2022," "Yes, in another venue," "Not sure," and "No." The test statistics also served as a measure of the effect sizes. *P* values were computed via permutation testing.

Results Overall, 322 papers were reviewed under the anonymous condition (116 accepted) and 310 papers under nonanonymous (114 accepted), with exactly 289 reviewers in both conditions. There were 611 discussion posts made by reviewers in the anonymous condition and 514 in the nonanonymous condition. The results for the 7 research questions are provided in **Table 24-0812**.

Conclusions Small but significant differences favoring anonymous discussions were found. Subsequent computer science conferences have drawn on these findings for their policy choices, with a greater inclination toward anonymity in reviewer discussions.

References

1. Shah N. An overview of challenges, experiments, and computational solutions in peer review (extended version).

Table 24-0812. Results for Each of the 7 Research Questions (RQs)

Research question	Sample size	Result
RQ 1	2281 Article-reviewer pairs	Marginally more discussion posts in the anonymous condition, with 611 posts on the 322 articles in the anonymous condition, and 514 posts on the 310 articles in the nonanonymous condition (*P* = .05)
RQ 2	484 Articles having both a senior and a junior reviewer	Decisions are closer to senior reviewers' scores in the nonanonymous condition than when anonymous (effect size, 0.15; *P* = .04)
RQ 3	1125 Discussion posts	No significant difference (Mann-Whitney normalized U statistic, 0.49; *P* = .72)
RQ 4	132 Respondents for each aspect	No significant difference for each of the 5 aspects queried in the survey (for each aspect, Mann-Whitney normalized U statistic between 0.43 and 0.48 and *P* > .30)
RQ 5	159 Respondents	Weak preference for anonymous discussions (Cohen *d* = 0.25)
RQ 6	159 Respondents	Most important: reviewers' feeling of safety in expressing their opinions (mean, 4.56); least important: politeness and professionalism of communication among reviewers (mean, 3.16)
RQ 7	167 Respondents	Approximately 7% reported having experienced such behavior either at UAI or another venue

July 7, 2025. Accessed July 16, 2025. https://www.cs.cmu.edu/~nihars/preprints/SurveyPeerReview.pdf

2. Chiang W, Li Z, Lin Z, et al. Vicuna: an open-source chatbot impressing GPT-4 with 90%* ChatGPT quality. LMSYSORG. March 2023. https://lmsys.org/blog/2023-03-30-vicuna/

[1]Google DeepMind, New York, NY, US; [2]Carnegie Mellon University, Pittsburgh, PA, US, nihars@cs.cmu.edu; [3]ETH Zurich, Zurich, Switzerland; [4]Max Planck Institute, Tübingen, Germany; [5]New Economic School, Moscow, Russia; [6]University of Maryland, Baltimore, MD, US.

Conflict of Interest Disclosures As per author affiliations above. In addition, Zhijing Jin is going to join the University of Toronto, and Kun Zhang has a partial appointment at Mohamed bin Zayed University of Artificial Intelligence.

Funding/Support ONR N000142212181, NSF 1942124, 2200410, 2229881, NIH R01HL159805.

Role of Funder/Sponsor The funders played 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 This work was conducted when Charvi Rastogi and Ivan Stelmakh were at Carnegie Mellon University. We thank James Cussens who served as program co-chair (with Kun Zhang), and general co-chairs Cassio de Campos and Marloes Maathuis of UAI 2022 for their support. We thank the OpenReview.net team, and particularly Harold Rubio, Melisa Bok, Celeste Martinez Gomez, and Nadia L'Bahy for helping us set up the experiment on the OpenReview.net platform. We also thank Giorgio Piatti for assistance to Zhijing Jin in setting up the large language models. We are grateful to all the participants of the UAI peer-review process for their time and effort. The experiment was reviewed and approved by the Carnegie Mellon University institutional review board.

Dual Anonymous and Distributed Peer Review for Proposal Review Rankings at the ALMA Observatory

John Carpenter,[1] Andrea Corvillón[1]

Objective In 2021,[1] the Atacama Large Millimeter/submillimeter Array (ALMA) transitioned from single-anonymous panel reviews to dual-anonymous distributed peer review to manage the growing volume of proposal submissions and mitigate potential biases.[2] We conducted a retrospective cohort study to examine associations between this procedural change and proposal rankings across principal investigator (PI) demographic characteristics, using 7 years of data under the previous format (2012-2018) and 4 years under the new process (2021-2024).

Design We analyzed proposal rankings from 12 ALMA cycles. From 2011 to 2018 (cycles 0-6), proposals were reviewed in topical panels under single-anonymous peer review. In 2019 (cycle 7), investigator lists were randomized while panels were retained. No review process was held in 2020 due to the COVID-19 pandemic. In 2021 (cycle 8), ALMA implemented dual-anonymous, distributed peer review for most proposals. We examined rankings by 3 PI demographic characteristics: (1) experience (number of cycles in which the PI submitted proposals), (2) regional affiliation (Chile, East Asia, Europe, North America, or other), and (3) sex. We grouped proposals by review era: single-anonymous panel review (cycles 1-6; 2012-2018; 9091 proposals) and dual-anonymous distributed peer review (cycles 8-11; 2021-2024; 6490 proposals). Cycle 0 and cycle 7 were excluded from the analysis: cycle 0 because all PIs were, by definition, first-time users of ALMA, and cycle 7 because it was a transitional year for implementing dual anonymity.

Results Proposal rankings were normalized from 0 (best) to 1 (worst) for comparability across cycles. Proposal counts by demographic subgroup and review era are reported in **Table 25-1053**. We compared median normalized rankings between review eras using 10,000 bootstrap samples to generate 95% CIs and 2-sided *P* values. We observed the following associations: (1) for experience, PIs submitting in all cycles had better rankings during single-anonymous panel review than under dual-anonymous distributed review ($P = .006$), and first-time PIs ranked lowest in both systems with no significant change in rankings ($P = .19$); (2) for regional affiliation, East Asian PIs showed improved rankings after the transition ($P < .001$), rankings for European PIs declined ($P = .009$) but remained above average, and rankings for PIs from North America, Chile, and other regions did not show significant changes ($P > .60$); and (3) for sex, no statistically significant differences in rankings were observed between male- and female-led proposals in either review system ($P = .12$).

Conclusions Dual-anonymous, distributed peer review was associated with reduced disparities by PI experience and region, consistent with reduced prestige and geographic bias. While increased score variability may have contributed, the nonuniform changes (ie, some groups improved, others remained stable) are inconsistent with a purely noise-driven explanation. These findings suggest that systematic shifts in reviewer behavior, not merely increased randomness, underlie the observed trends. Disentangling the effects of dual anonymity, distributed review, and other concurrent changes (eg, increased use of artificial intelligence) remains an important direction for future research.

References

1. Donovan Meyer J, Corvillón A, Carpenter JM, et al. Analysis of the ALMA Cycle 8 distributed peer review process. *Bull Am Astron Soc*. 2022;54(1):43. doi:10.3847/25c2cfeb.4ece85d4

2. Carpenter JM, Corvillón A, Donovan Meyer J, et al. Update on the systematics in the ALMA proposal review process after Cycle 8. *Publ Astron Soc Pac*. 2022;134:045001. doi:10.1088/1538-3873/ac5b89

[1]Joint ALMA Observatory, Santiago, Chile, john.carpenter@alma.cl.

Conflict of Interest Disclosures John Carpenter and Andrea Corvillón are employed by the Joint ALMA Observatory, which is jointly managed by Associated Universities Inc/National Radio Astronomy Observatory, the European Organisation for Astronomical Research in the Southern Hemisphere, and the

Table 25-1053. Proposals and Review Rankings by Demographic Characteristic

Characteristic	Proposals, No.		Review rankings, median[a]		Ranking change, difference (95% CI)[b]	*P* value
	Before	After	Before	After		
Experience						
All-cycle PIs[c]	3211	743	0.414	0.467	−0.053 (−0.099 to −0.009)	.006
First-time PIs	1953	860	0.621	0.647	−0.026 (−0.069 to 0.010)	.19
Region						
Chile	565	344	0.622	0.607	0.015 (−0.046 to 0.099)	.61
East Asia	1810	1495	0.625	0.562	0.063 (0.033 to 0.092)	<.001
Europe	3815	2526	0.457	0.487	−0.030 (−0.053 to −0.008)	.009
North America	2642	1887	0.435	0.429	0.007 (−0.020 to 0.039)	.60
Other	259	238	0.682	0.676	0.006 (−0.083 to 0.067)	.94
Sex						
Male	6202	4178	0.497	0.488	0.009 (−0.012 to 0.030)	.40
Female	2889	2308	0.506	0.518	−0.012 (−0.040 to 0.015)	.42
Difference, male – female	NA	NA	−0.009	−0.030	0.021 (−0.013 to 0.055)	.12

Abbreviations: NA, not applicable; PIs, principal investigators.

[a]Proposal rankings were normalized from 0 (best) to 1 (worst).

[b]Calculated as ranking before minus ranking after. A positive difference signifies that rankings improved.

[c]All-cycle PIs are those who submitted in every cycle up to and including the one analyzed (eg, cycles 0-2 for cycle 2)

National Astronomical Observatory of Japan on behalf of the ALMA partnership.

Funding/Support This work was supported by the Joint ALMA Observatory.

Role of the Funder/Sponsor The funder had a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and decision to submit the abstract for presentation.

Acknowledgment ALMA is a partnership of the European Organization for Astronomical Research in the Southern Hemisphere (representing its member states), National Science Foundation (US), and National Institutes of Natural Sciences (Japan), together with the National Research Council of Canada (Canada), the National Science and Technology Council and Academia Sinica Institute of Astronomy and Astrophysics (Taiwan), and the Korea Astronomy and Space Science Institute (Republic of Korea), in cooperation with the Republic of Chile.

Comparison of Content in Published and Unpublished Peer Review Reports

Elena Álvarez-García,[1] Daniel Garcia-Costa,[1] Flaminio Squazzoni,[2] Mario Malički,[3,4,5] Bahar Mehmani,[6] Francisco Grimaldo[1]

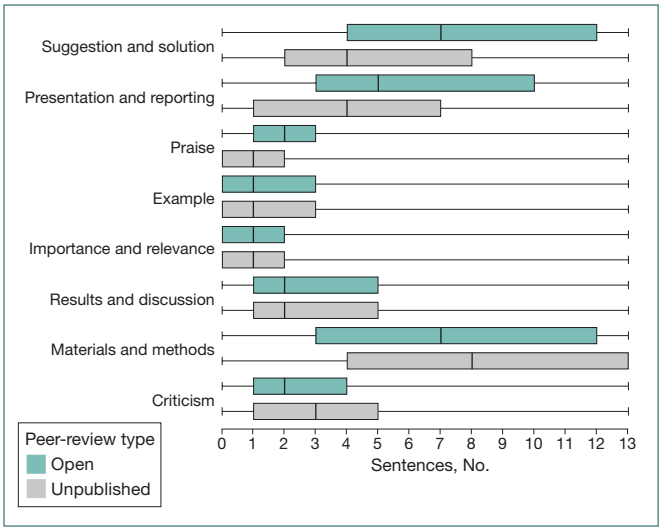
Objective While the publication of review reports increases the transparency of peer review, little is known about the effect it has on the content or quality of reports.[1] The objective of our study was to assess differences in length, information content, and similarity of open vs unpublished reports.

Design This was a cross-sectional study following STROBE guidelines for reporting[2] that compared open with unpublished reports from 233 medical journals from Elsevier and Springer Nature, submitted from 2016 to 2021. Reports were obtained through confidential agreements with the publishers. Content of reports was compared using number of sentences, previously validated models for classification of 8 categories of content,[3] information content score (mean cumulative distribution function values of each category's sentence count, based on a zipfian distribution, with scores ranging from 0 to 1, where 1 indicates the highest informativeness), and information score dispersion measured by the Gini index (measure of similarity between reports for the same manuscript, where 0 indicates total similarity and 1 indicates total dissimilarity). Number of sentences, information score, and dispersion were compared using Mann-Whitney *U* tests. The association of information score and Gini index with peer review type, journal quartile of impact factor, and reviewer gender, seniority, and region were explored using generalized linear models. To adjust for multiple comparisons, we considered $P < .001$ as statistically significant.

Results Compared with unpublished reports (n = 117,250), open peer review reports (n = 40,844) were longer (median [IQR] length, 22 [12-37] sentences vs 17 [10-28] sentences; $P < .001$) than unpublished reports and had a higher informative content (median [IQR] score, 0.52 [0.35-0.70] vs 0.46 [0.30-0.65]; $P < .001$), with the largest difference found in the number of suggestion and solution sentences (**Figure 25-1081**). Women's reports had a higher information score than men's reports (difference, 6.3%), and reviewers from non-Western institutions had lower scores than those from Western institutions (difference, −6.0%). Open peer review reports were also more similar to each other (median [IQR] Gini index, 0.19 [0.09-0.33] vs 0.22 [0.11-0.35]; $P < .001$).

Conclusions Our study showed that open peer review reports were longer than traditional unpublished reports, with the greatest differences found in the number of suggestion and solution sentences. These results suggest that increasing the transparency of peer review could lead to more

Figures 25-1081. Box Plots of Number of Sentences per Category in Open vs Unpublished Peer-Review Reports



Midlines indicate medians; boxes, IQRs; and whiskers, ranges.

detailed reports that focus on manuscript improvement. A limitation of our study is that we did not have access to published manuscripts and were unable to determine the impact of initial quality of manuscripts beyond journal ranking, nor the impact peer review reports had on manuscript improvement.

References

1. Ross-Hellauer T, Horbach SP. Additional experiments required: a scoping review of recent evidence on key aspects of open peer review. *Res Eva*. Published online February 8, 2024. doi:10.1093/reseval/rvae004

2. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495-1499. doi:10.1016/j.ijsu.2014.07.013

3. Severin A, Strinzel M, Egger M, et al. Relationship between journal impact factor and the thoroughness and helpfulness of peer reviews. *PLoS Biol*. 2023;21(8):e3002238. doi:10.1371/journal.pbio.3002238

[1]Department of Computer Science, University of Valencia, Valencia, Spain; [2]Department of Social and Political Sciences, University of Milan, Milan, Italy; [3]Stanford Program on Research Rigor and Reproducibility, Stanford University, Stanford, CA, US, mmalicki@stanford.edu; [4]Department of Epidemiology and Population Health, Stanford University, Stanford, CA, US; [5]Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, US; [6]STM Journals, Elsevier, Amsterdam, the Netherlands.

Conflict of Interest Disclosures Mario Malički was an Editor in Chief of *Research Integrity and Peer Review*, which is published by Springer Nature, provider of the part of the dataset. Bahar Mehmani is an employee of Elsevier, provider of the other part of the dataset. Bahar Mehmani 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 Elena Álvarez-García, Daniel Garcia-Costa, and Francisco Grimaldo have been partially supported by the Regional Ministry of Education, Culture, Universities and Employment of the Generalitat Valenciana under project CIAICO/2022/154. Flaminio Squazzoni was supported by a grant from the Progetti di Rilevante Interesse Nazionale—Italian Ministry of University and Research (grant 202297CKET_00 "ALGOLIT").

Role of the Funder/Sponsor The funder was 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; and decision to submit the abstract for presentation.

Acknowledgment We gratefully acknowledge the support of Springer Nature, which provided the open peer review reports for our study, specifically Suzuki Limbu, Attilia Czikmantori, and their production team. We also thank all reviewers of Springer Nature journals that participated in signing or allowing their review reports to be published and all journals and editors facilitating these processes. We acknowledge the support on data extraction from the IT staff of Elsevier, specifically Ramsundhar Baskaravelu and his team. This work uses Scopus data provided by Elsevier through the Peer Review Workbench. We also thank Dave Santucci from the Elsevier Scopus API team and Kristy James from the International Center for the Study of Research for their support on data enrichment about authors and reviewers. We thank Josep Monclús for his help on data extraction and preparation during the initial phase of this study.

Editorial and Publishing Processes and Models

Changes to Research Article Abstracts Between Submission and Publication

Christos P. Kotanidis,[1,2,3] Sarah Gorey,[1,4] Harleen Marwah,[1] Abarna Pearl,[1] Darren Taichman,[1] Mary Beth Hamel[1,5]

Objective Editorial review forms the cornerstone of scientific publishing, ensuring that published research adheres to the highest quality standards.[1] We evaluated how editorial processes shape manuscripts by examining changes in abstracts between submission and publication.

Design We assessed research articles submitted to the *New England Journal of Medicine* (*NEJM*) in 2022 and published in *NEJM* or rejected and subsequently published in 437 other journals. Using author names, we searched PubMed for articles within 2 years of submission. Four physicians reviewed search results to match published manuscripts with submitted manuscripts. Abstract similarity was quantified using term frequency–inverse document frequency,[2,3] assigning a cosine similarity score ranging from 0 (dissimilar) to 1 (identical). To assess if similarity changes reflected substantive differences, masked physician reviewers conducted qualitative comparisons of submitted and published versions of abstracts from clinical trials. Abstracts were assessed for substantive changes in the following 4 domains: trial design, primary outcome, adverse events, and conclusion (TPAC). A score of +1 was assigned for a substantive domain change favoring the published version and −1 for domain change favoring the submitted version,

with the domain scores summed, resulting in a total TPAC score range of −4 (submitted better) to +4 (published better).

Results We matched 1360 PubMed records from 2756 research articles initially submitted to *NEJM*. Of these, 201 were published in *NEJM* and 1159 were rejected by *NEJM* and published in the 437 other journals. The median (IQR) similarity score between submitted and published abstracts was 0.71 (0.58-0.83). Abstracts of articles published in the 5 general medicine journals with a journal impact factor (JIF) above 50 per the 2023 Clarivate report had significantly lower similarity scores compared with those published in other journals (**Figure 25-0940**, A). For qualitative comparisons, we identified abstracts from randomized controlled trials. On average, published abstracts improved in 0.9 domains, with conclusion being the most frequently modified domain (44.2%). Similarity score correlated negatively with TPAC (Spearman ρ coefficient, −0.39; **Figure 25-0940**, B), indicating that lower similarity reflected more substantive revisions. Among abstracts published in the 5 high-JIF general medicine journals, 72.1% improved in at least 1 domain, compared with 48.3% of abstracts published in the other 433 journals. This difference was consistent across all TPAC domains, with conclusion changes being the most different (48% vs 27%).

Conclusions Our comparisons of submitted and published abstracts suggest that peer review and editorial processes result in substantial revision of research reports. This report focused on submissions to 1 general medical journal and within a 2-year publication window. These findings reinforce the importance of editorial oversight in scientific publishing, highlighting its role in refining study design communication, reporting of outcomes, and overall clarity in research dissemination.

References

1. Roll SC. The value and process of high-quality peer review in scientific professional journals. *J Diagn Med Sonography*. 2019;35(5):359-362. doi:10.1177/8756479319853800

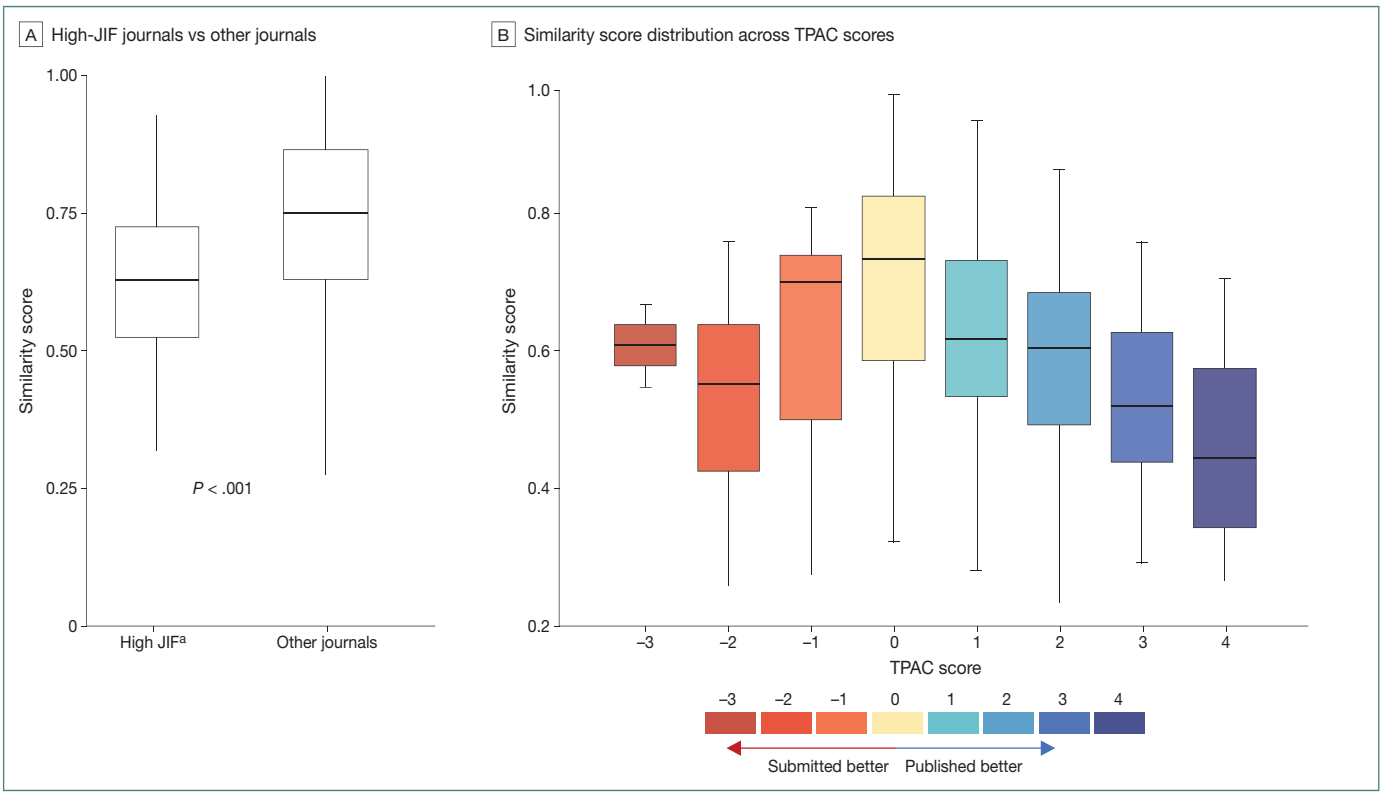
2. Aizawa A. An information-theoretic perspective of tf–idf measures. *Inf Process & Manage*. 2003;39(1):45-65. doi:10.1016/S0306-4573(02)00021-3

3. Ramos J. Using tf-idf to determine word relevance in document queries. In: *Proceedings of the First Instructional Conference on Machine Learning*. 2003.

[1]*New England Journal of Medicine*, Boston, MA, US, ckotanidis@nejm.org; [2]Heart and Vascular Center, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US; [3]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; [4]Stroke Clinical Trials Network Ireland (SCTNI), University College Dublin Clinical Research Centre, Mater Hospital, Dublin, Ireland; [5]Division of General Medicine and Primary Care, Department of Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA, US.

Conflict of Interest Disclosures Christos P. Kotanidis, Sarah Gorey, Harleen Marwah, and Abarna Pearl are Editorial Fellows, Darren Taichman is a Deputy Editor, and Mary Beth Hamel is the Executive Editor at the *New England Journal of Medicine*.

Figure 25-0940. Abstract Similarity in High Impact Factor vs Other Journals and Across TPAC Scores



[a]High journal impact factor (JIF) general medicine journals: *Lancet, NEJM, BMJ, JAMA, Nature Medicine.*

Manuscript Characteristics Associated With Editorial Review and Peer Review Outcomes at *Science* and Science *Advances*

Nicholas LaBerge,[1] Sam Zhang,[1,2,3] Quinten McElhiney,[1] Daniel B. Larremore,[1,2] Aaron Clauset[1,2]

Objective To make publicly available a deidentified dataset of manuscript submissions and associated editorial metadata at *Science* and *Science Advances* (2 elite multidisciplinary journals) and quantify the manuscript and author characteristics associated with outcomes over the 2 stages of evaluation at these highly selective journals: (1) editorial review, when journal editors screen and select a smaller set of submissions for detailed consideration, and (2) peer review, when editors recruit outside experts to evaluate manuscripts. Peer review at elite scientific journals is a high-stakes process whose outcomes have a broad influence in science and society. However, the need to maintain peer review's confidentiality has limited the range of data available for scientific study. This lack of peer review data makes it difficult to assess how well elite journals achieve the scientific community's ideals or to improve and test theories to improve their evaluation processes.

Design We introduced and described a manuscript-level dataset of 110,303 deidentified evaluations of manuscripts submitted to *Science* and *Science Advances* over a 5-year period (2015-2019), each with a standard set of author and manuscript characteristics, and we conducted a series of logistic regression analyses to quantify the correlates of success in the initial editorial review stage (desk rejections) and the subsequent peer review stage at both journals.

Results Each manuscript record includes author, editor, and manuscript characteristics, including topic, team size, institutional prestige, gender, geographic region, evaluation scores, and editor decisions; personally identifiable and reidentifiable information, including the text of the reviews, is excluded. Our statistical analyses revealed strong associations with institutional prestige, team size, manuscript topic, and country, which are primarily attributable (via a mediation analysis) to the influence of the editor, even as the tenor of advice from outside experts correlates strongly with the final editorial decision (**Figure 25-0972**). Corresponding authors who are men appear to have a small but significant advantage with editors and reviewers at *Science*, while authors based in China have a significant disadvantage.

Conclusions This deidentified dataset will support further investigation by peer review researchers into the 2-stage evaluation process of these 2 elite journals, and our results will help quantify the complex interactions among editors, reviewers, manuscript characteristics, and author characteristics at elite journals.

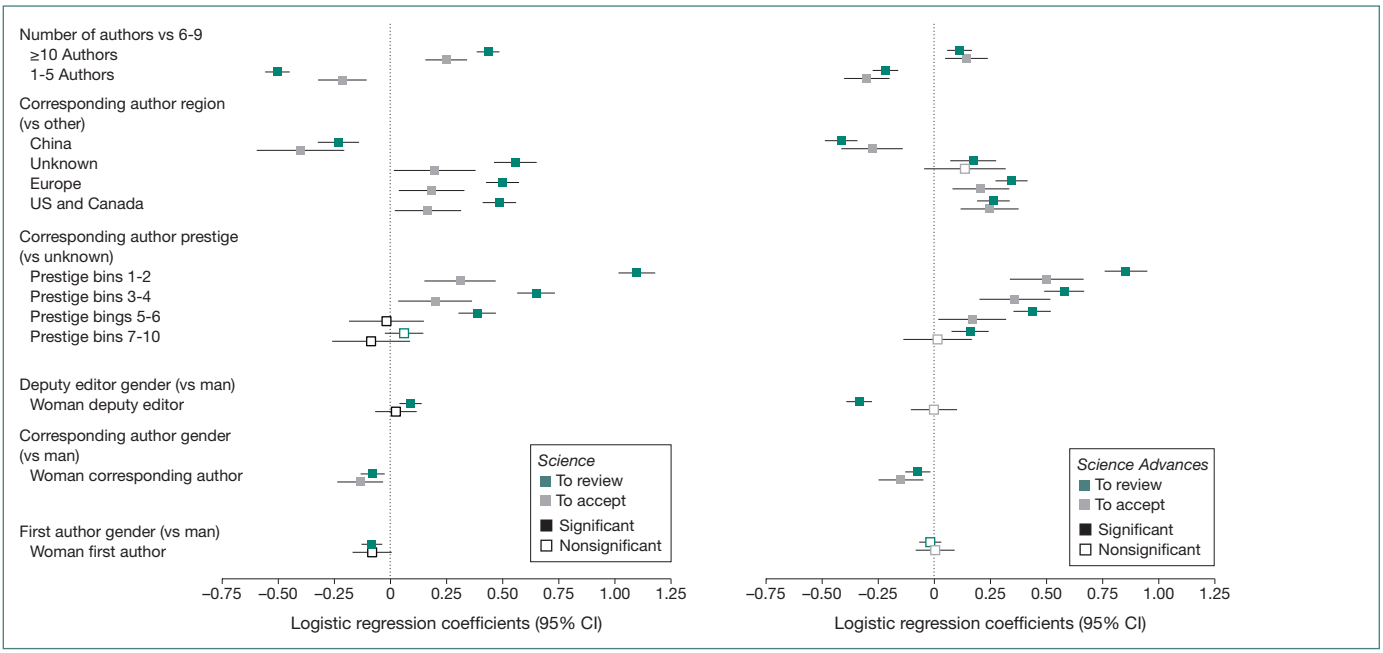
[1]University of Colorado Boulder, Boulder, CO, US, aaron.clauset@colorado.edu; [2]Santa Fe Institute, Santa Fe, NM, US; [3]University of Vermont, Burlington, VT, US.

Conflict of Interest Disclosures Aaron Clauset is a Deputy Editor at *Science Advances*; all other authors declare no conflicts of interest.

Funding/Support This work was supported by Air Force Office of Scientific Research Award FA9550-19-1-0329 (to Nicholas LaBerge, Sam Zhang, Daniel B. Larremore, and Aaron Clauset), National Science Foundation (NSF) SBE Award 2219609 (to Nicholas LaBerge, Sam Zhang, Daniel B. Larremore, and Aaron Clauset), NSF Graduate Research Fellowship Award DGE 2040434 (to Sam Zhang), and NSF Alan T. Waterman Award SMA-2226343 (to Daniel B. Larremore).

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

Figure 25-0972. Manuscript Characteristics and Review Stage Success



Logistic regression coefficients for each nontopic manuscript covariate at *Science* (left) and *Science Advances* (right), adjusting for topic, for 110 303 submissions between 2015 and 2019: (green) progressing from editorial review to peer review and (gray) being accepted for publication after peer review. Reference categories are shown in brackets in each covariate label; statistical significance is indicated with a solid marker (*z* score test, α = .05).

Investigating Changes in Common Vocabulary Terms in *eLife* Assessments Across Versions, in a Publish, Review, Curate Model

Nicola Adamson,[1] Andy Collings[1]

Objective In October 2022, *eLife* announced a Publish, Review, Curate model[1] in which all peer-reviewed submissions are published as reviewed preprints, accompanied by an *eLife* Assessment[2] and public reviews. Authors decide which revisions to undertake and when to declare a version of record (VOR), without an accept/reject decision after peer review. We investigated the extent to which authors revise and improve their work.

Design An *eLife* Assessment is prepared for each submission sent for review and subsequent revisions, which summarizes the strength of evidence and significance of findings using terms from a common vocabulary. We extracted terms used in *eLife* Assessments for all 2918 reviewed preprints and their subsequent versions to the end of 2024. To ensure publications had completed the entire process and we could evaluate the full extent of changes, we selected the 1504 where a VOR had been declared. From these, 213 were excluded because either the VOR or original reviewed preprint had included multiple terms to describe the evidence. Using a retrospective analysis, we studied the remaining 1291 reviewed preprints and their respective VOR, calculating the distributions of and changes to terms between first and final versions and the number of rounds of revision.

Results Panel A in **Figure 25-0989** shows the terms used to describe the strength of evidence in VORs and the first version of the reviewed preprints. The evidence term improved between versions in 39.4% of cases (n = 509) and remained the same in 57.7% of VORs (n = 745). Of VORs in which the evidence was described as incomplete in the first reviewed preprint, 76.5% (n = 199) improved to solid or better, and in papers originally described as solid, 49.3% improved to convincing or better (n = 217). In papers in which the evidence was originally described as inadequate (n = 23), this improved in 78.3% of instances (n = 18) and approximately one-half ended as solid or better (47.8%, n = 11). The majority of VORs (82.6%, n = 1066) were declared after 1 round of revision and 13.5% (n = 174) after 2 or more rounds. Of the 4.0% of VORs declared without revisions (n = 51), only 13.7% had an evidence term of inadequate (n = 0) or incomplete (n = 7). The significance of the findings (**Figure 25-0989**, B) was most frequently described using valuable (VOR: 31.1%, n = 401) or important (VOR: 49.0%, n = 633); overall, significance terms remained the same in 78.2% of cases (n = 1010).

Conclusions Authors generally complete at least 1 round of revision, even though they have the option to proceed without, and multiple rounds of revision are uncommon. Where the evidence was originally described as incomplete, the majority of authors revised to improve this before the VOR. Terms used to describe the significance of the findings change less often than evidence terms, perhaps due to significance being more closely linked to the original research question under investigation.

References

1. Eisen MB, Akhmanova A, Behrens TE, et al. Scientific publishing: peer review without gatekeeping. *eLife*. October 20, 2022;11:e83889. doi:10.7554/eLife.83889

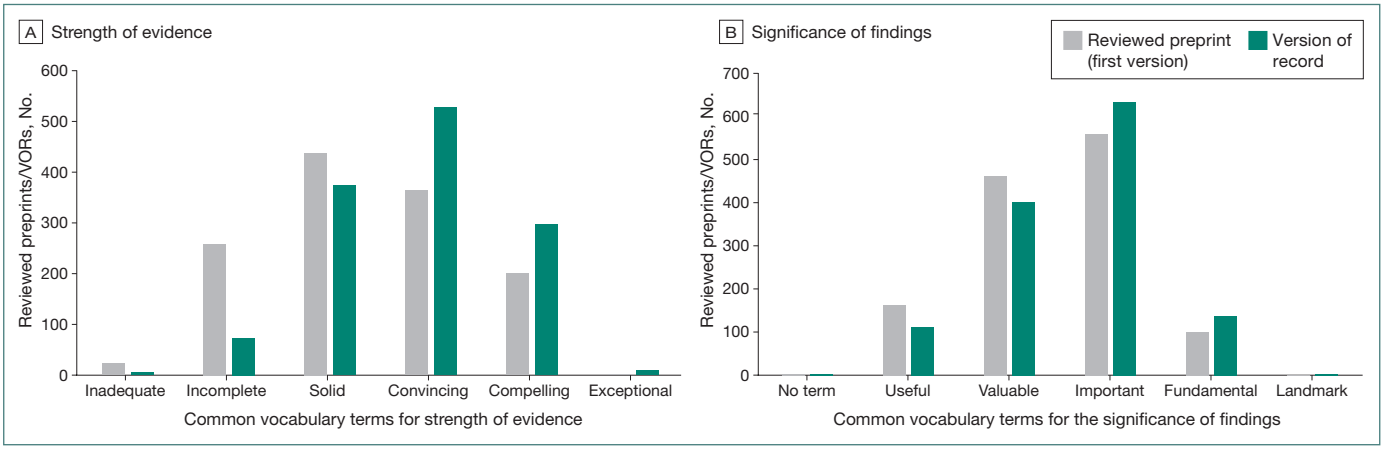
2. *eLife*'s New Model: What is an *eLife* assessment? *eLife*. October 20, 2022. Accessed January 30, 2025. https://elifesciences.org/inside-elife/db24dd46/elife-s-new-model-what-is-an-elife-assessment

[1]eLife Sciences Publications Ltd, 95 Regent Street, Cambridge, CB2 1AW, UK, n.adamson@elifesciences.org.

Conflict of Interest Disclosures We have no competing interests to declare. Both authors are employed by eLife Sciences Publications Ltd; we received no external funding for this work.

Acknowledgment We thank Fiona Hutton, Fred Atherden, Emily Packer, and George Currie for their comments and suggestions on the submission, and Fred Atherden for extracting data for analysis.

Figure 25-0989. Terms From Common Vocabulary for Strength of Evidence and Significance of Findings



Frequency of each term used to describe (A) the strength of evidence and (B) the significance of findings in the *eLife* Assessment for *eLife* model versions of record (VORs) to the end of 2024 and the respective first version reviewed preprint.

Peer Review Times and Payment Incentives

Results of Testing the Gold Standard 2-Week Reviewer Deadline

Emilie Gunn,[1] Kelly Brooks,[1] Stephanie Valladao[1]

Objective Prior research has shown the effects of shortening a review deadline on overall time to review.[1] We wanted to determine whether extending the standard 2-week deadline for peer review in a scholarly journal would have positive outcomes. In particular, we wanted to test the hypothesis that offering a 21-day deadline would increase the reviewer conversion rate, thus reducing the number of reviewers invited before getting 2 reviewers to agree. A secondary end point was the number of days it took reviewers to complete their reviews.

Design In this pilot study, 593 review invitations for 95 manuscripts were sent from *JCO Oncology Practice* between July 1, 2024, and September 30, 2024, with a 21-day deadline. This was compared with 606 invitations for 116 manuscripts sent between July 1, 2023, and September 30, 2023, with a 14-day deadline. Only Original Reports and Reviews were studied because these article types require at least 2 reviewers. Using the Editorial Manager submission system, editors invited reviewers using the normal process, except that reviewers were offered 21 days instead of the industry-standard 14 to complete their reviews. Reviewers received information regarding the 21-day time frame at initial invitation, at the time of review invitation acceptance, and via reminders 1 week and 2 days prior to the deadline. The goal for each paper was to obtain 2 reviewers.

Results When the time allowed to submit a review was increased to 21 days, the number of reviewers invited per manuscript increased by 0.5 reviewers (10.4%). The time it took reviewers to accept a review invitation increased by 0.1 days (12.5%), measured as the number of days from the date the invitation was sent to the date the reviewer agreed (**Table 25-0847**). The percentage of late reviews decreased by 5% (from 27% to 22%). The median (IQR) time to submit a review was 19 (11-33) days, compared with 13 (5-18) days during the control period. Additionally, late reviews were submitted an average of 11.1 days late, a 29.1% increase. However, median (IQR) time from submission to first decision decreased by 4.7% to 41 (31-55) days from 43 (30-65.5) days.

Conclusions Extending the 2-week deadline for peer review did not decrease the number of reviewers invited before getting 2 to agree. When the allowed time was increased, the time taken to review increased as well. Although there were fewer late reviews overall, the reviews submitted past the deadline were later than those submitted under the 14-day deadline. We ultimately recommended that the journal continue with the current 2-week deadline for reviewers. Further research is needed in this area.

Table 25-0847. Summary of Findings

	Under 14-day deadline	Under 21-day deadline	Actual change	Percentage change
No. of manuscripts	116	95	–21	–22.1
No. of review invitations sent	606	593	–13	–2.2
Average No. of reviewers invited per manuscript	4.8	5.3	0.5	10.4
Average No. of days to accept invite	0.8	0.9	0.1	12.5
Median (IQR) No. of days to submit a review	13 (5-18)	19 (11-33)	6	31.6
Average No. of days late to submit a review	8.6	11.1	2.5	29.1
Median (IQR) No. of days from submission to first decision	43 (30-65.5)	41 (31-55)	–2	–4.7

Reference

1. Cochran A, Guertin L. Affecting editor behavior with data: a case study. *Sci Editor*. 2014;37(1):117-119.

[1]American Society of Clinical Oncology, Alexandria, VA, US, emilie.gunn@asco.org.

Conflict of Interest Disclosures None reported.

Acknowledgment We thank Jeffrey Peppercorn for his help with designing this abstract and Scott Marchese for his statistical expertise.

Analysis of Decisions and Lead-Time in Ethical Review Boards in Sweden

Emmanuel A. Zavalis,[1] Love Ahnström,[2] Natasha Olsson,[2] Gustav Nilsonne[2,3]

Objective The objective of this study was to assess lead times and application decisions using the national database of the Swedish Ethical Review Authority since its inception. We also aimed to assess the effect of the COVID-19 ethical review fast-track implemented in spring 2020. Prior studies of ethical review have been limited to single or few institutions.1,2

Design This retrospective analysis examined 21,962 administrative records of all ethical review applications submitted in Sweden between 2021 and 2023. The applications were categorized by type (eg, amendments, clinical trial, or processing sensitive data) and sponsor (eg, health care systems, universities, or corporations). Kaplan-Meier curves were constructed for analysis of time to decision. Adherence to statutory deadlines was assessed for amendments (35 days) and drug trials (60 days).3 Log-rank test was used for statistical testing with $\alpha = .05$ as a significance threshold.

Results Following exclusions (eg, incomplete applications, unpaid fees, or withdrawals), 20,440 applications remained: 5866 from 2021, 6926 from 2022, and 7648 from 2023. Submissions were primarily from health care systems (10,456

[51%]) and universities (8681 [42%]), with corporations accounting for 578 applications (3%). Amendment applications constituted the largest category of applications (51% in 2021, 50% in 2022, and 44% in 2023). Applications in the "processing of personal data" category doubled from 736 to 1574 over the study period (**Figure 25-0966,** A). Of the 20,440 applications, initial approvals accounted for 68% (n = 13,864), further information or change to the project was requested in 24% (n = 5001) of cases, and 1.5% (n = 313) were rejected. Final decisions resulted in approvals for 91% (n = 18,516), with rejections being rare (2%), as seen in Figure 25-0966, B. Geographic variation was observed, with Stockholm having the shortest (34 days) and Lund the longest (43 days) median lead times. COVID-19 studies were associated with expedited processing times (median time to decision, 51 vs 63 days; P < .001). While 27% of amendment applications were without a decision after 35 days, 39% of original applications as well as 58% of drug trials were without a decision after 60 days (Figure 25-0966, C).

Conclusions The Swedish Ethical Review Authority mainly handled amendments. Lead times were longer than statutory deadlines for more than half of drug trials and for a quarter of amendments. The processing times showed geographic variations, and the fast-track for COVID-19 showed a significant association with processing times. Further studies of the qualitative aspects of ethical reviews' impact on research plans as well as the nature of the applications for amendments would be of interest for a better understanding of the value introduced and the harms prevented by ethical review.

References

1. Hall DE, Hanusa BH, Stone RA, Ling BS, Arnold RM. Time required for institutional review board review at one Veterans Affairs medical center. *JAMA Surg*. 2015;150(2):103. doi:10.1001/jamasurg.2014.956

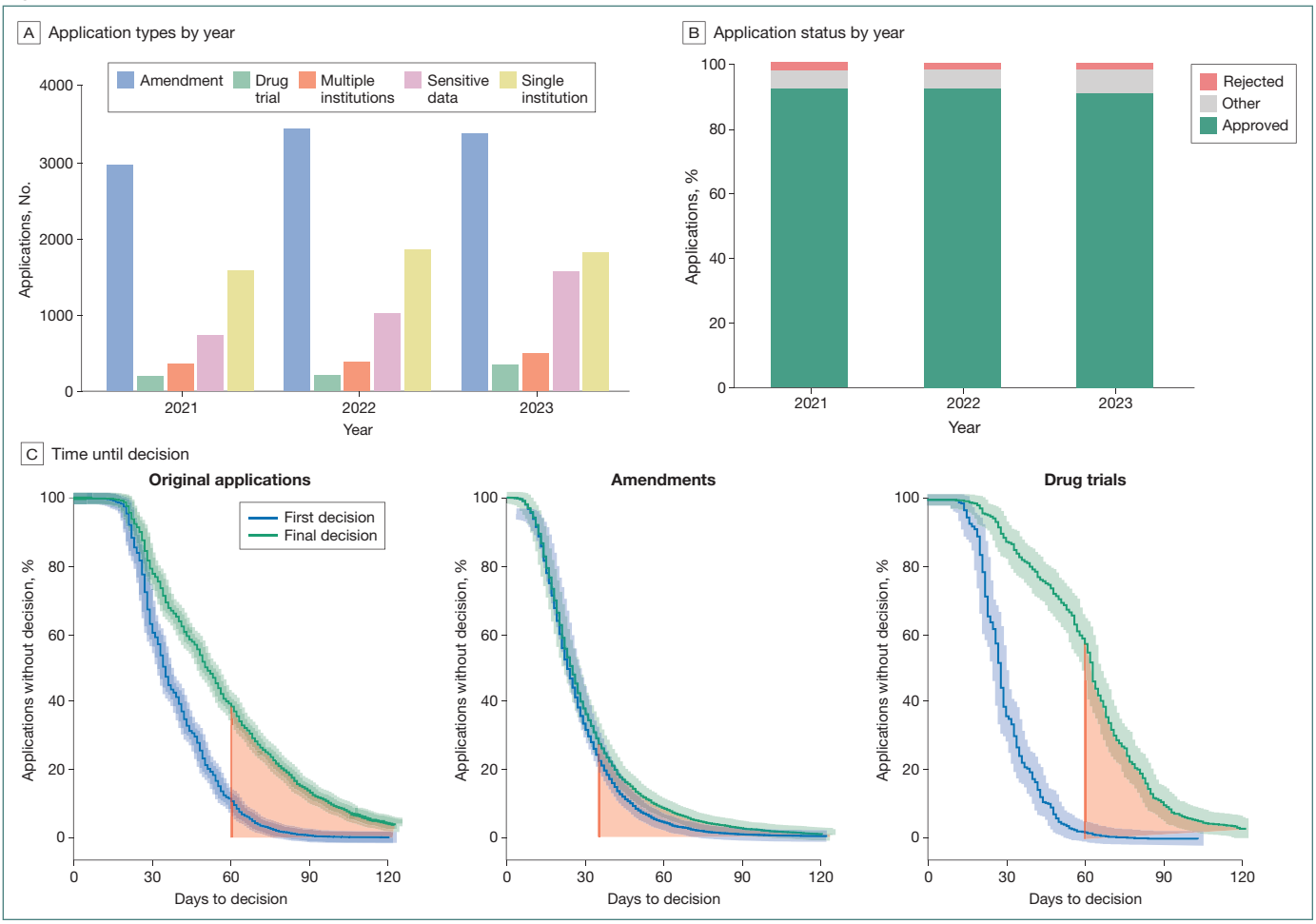
2. Varley PR, Feske U, Gao S, et al. Time required to review research protocols at 10 Veterans Affairs institutional review boards. *J Surg Res*. 2016;204(2):481-489. doi:10.1016/j.jss.2016.06.004

3. Sveriges Riksdag. Ordinance (2003:615) on ethical review of research involving humans. Accessed January 18, 2025. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/forordning-2003615-om-etikprovning-av-forskning_sfs-2003-615/

[1]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden; [2]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden, gustav.nilsonne@ki.se; [3]Department of Psychology, Stockholm University, Stockholm, Sweden.

Conflict of Interest Disclosures None reported.

Figure 25-0966. Lead Times and Decisions in Swedish Ethical Review



Funding/Support This project was funded by Sweden's Innovation Agency (Vinnova).

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

Acknowledgment We thank Iris Rocamonde Lago, who helped with the visualizations.

Monetary Incentives for Peer Review at a Medical Journal: A Quasi-Randomized Experimental Study

Christopher S. Cotton,[1] Abid Alam,[1] Sophie Tosta,[2] Timothy G. Buchman,[3] David M. Maslove[4,5,6]

Objective Peer review is a cornerstone of academic publishing, providing a bulwark against the dissemination of flawed or unethical research. Historically, peer review has been carried out by experts on a volunteer basis. Financial compensation for peer reviewers has been discussed but with little empirical evidence to inform the debate.[1] A single experimental study[2] has examined this question, finding that for peer review at an economics journal, offering a $100 gift card to reviewers reduced median review time by 8 days. No experimental study to our knowledge has looked at the effect of paying for peer review at a medical journal. Our objective was to determine the effect of a cash incentive on the rate of fulfillment of peer review requests at a medical journal.

Design We conducted a quasi-randomized experimental study at *Critical Care Medicine*, a specialty medical journal. We divided the experimental period (September 2023 to March 2024) into alternating 2-week blocks, during which all peer review requests sent out to prospective reviewers included either an offer of $250 to complete the peer review assignment (incentive blocks) or no offer (control blocks). Only original research reports were included in the study, with the monetary incentive applying only to the initial review. The primary outcome was the rate of invitation –to completed review conversion, defined as the ratio of reviews submitted divided by the number of reviewer invitations sent out. Review quality was a secondary outcome.

Results During the study, 131 manuscripts went out for review, with a median (IQR) of 5 (2-8) reviewer invitations per manuscript. Of 715 reviewer invitations, 414 (58%) included an incentive offer. There was no difference in the proportion of invitations that were accepted (control vs incentive, 144 of 301 [48%] vs 218 of 414 [53%]; $P = .20$), but a difference was seen in the primary outcome, with reviewers in the intervention group having a greater odds of completing the review compared with those in the control group (control vs incentive, 127 of 301 [42%] vs 206 of 414 [50%]; odds ratio, 1.36; 95% CI. 1.01-1.84). A Cox proportional hazards model showed that incentivized invitations were fulfilled slightly faster (median [IQR] time, 11 [6-14] days vs 12 [7-14] days; hazard ratio, 1.30; 95% CI, 1.04-1.62). There was no difference in the quality of reviewer reports, as adjudicated by the handling editors (control vs incentive, median [IQR] score on a 100-point scale, 75 [70-85] vs 75 [66-90]).

Conclusions Our findings suggest that cash payments can incentivize peer review at a medical journal, potentially providing a method to help journals expedite peer review in situations like public health emergencies. Important policy questions remain around cost-benefit considerations; extrapolating from the rate of reviews required during the study period, we estimate that offering $250 for each review would cost *Critical Care Medicine* nearly $150,000 annually. Other considerations include sources of funding for paid reviews, the effect of nonmonetary incentives, and the potential unintended consequences of paid reviews.

References

1. Cheah PY, Piasecki J. Should peer reviewers be paid to review academic papers? *Lancet*. 2022;399(10335):P1601. doi: 10.1016/S0140-6736(21)02804-X

2. Chetty R, Saez E, Sándor L. What policies increase prosocial behavior? an experiment with referees at the Journal of Public Economics. *J Econ Perspect*. 2104;28(3):169-188. doi: 10.1257/jep.28.3.169

1Department of Economics, Queen's University, Kingston, ON, Canada; [2]Society of Critical Care Medicine, Mount Prospect, IL, US; [3]Emory Critical Care Center, Emory University School of Medicine, Atlanta, GA, US; [4]Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada, david.maslove@queensu.ca; [5]Department of Medicine, Queen's University, Kingston, ON, Canada, [6]Kingston Health Sciences Centre, Kingston, ON, Canada. Department of Medicine, Queen's University, Kingston, ON, Canada, 6Kingston Health Sciences Centre, Kingston, ON, Canada.

Conflict of Interest Disclosures David M. Maslove was an associate editor at *Critical Care Medicine* at the time the study was conducted. Timothy G. Buchman was the editor-in-chief at *Critical Care Medicine* at the time the study was conducted. Sophie Tosta is the managing editor of *Critical Care Medicine*.

Funding/Support This study was supported by a grant from the Government of Canada's New Frontiers in Research Fund (grant No. NFRFR-2021-00335). David M. Maslove is supported by a New Clinician Scientist Award from the Southeastern Ontario Academic Medical Association. Christopher S. Cotton receives research funding as the Jarislowsky-Deutsch Chair in Economic & Financial Policy at Queen's University.

Role of Funder/Sponsor No funder had a role in the design, analysis, or interpretation of the study.

Acknowledgment The authors wish to thank the associate editors and senior editors at *Critical Care Medicine*, as well as the editorial assistants at the journal, and the staff at the Society of Critical Care Medicine for facilitating the study.

Exploring Views on Remuneration for Review: A Survey of *BMJ*'s Patient and Public Reviewers

Sara Schroter,[1,2] Rebecca Harmston,[3] Emma Doble,[1] Sophie Cook,[1] Amy Price[4,5]

Objective Calls for patient partners to be paid for their contributions to the health sector are growing.[1] *The BMJ* invites patients and the public, as well as academic and

clinical peer reviewers, to review manuscripts.[2] On November 28, 2024, *BMJ* announced it would remunerate its patient and public reviewers from 2025.[3] We surveyed these reviewers to capture their perspectives on the introduction of remuneration and overall experience of reviewing.

Design Survey of current patient and public reviewers (those who had completed a review in the previous 3 years) on SurveyMonkey on December 11, 2024. We asked reviewers about their experiences of reviewing and their perspectives on the introduction of remuneration (£50 or an online subscription to a BMJ journal was provided for each completed review).

Results We received a response from 183 of 267 invited reviewers (69%). Independent *t* tests showed that respondents, compared with nonresponders, received more invitations (mean: 14.0 vs 10.3; $P = .01$) and completed more reviews (mean: 7.7 vs 3.7; $P < .001$), but there was no difference in 5-point review quality score (mean: 3.8 vs 3.8; $P = .81$). The majority of respondents were based in the UK and the US. Respondents were positive about their experience of reviewing: 84% (n = 154) reported it as a very good or good experience. The primary motivation for 93% of reviewers (n = 170) was contributing patient perspectives to the research process. Regarding remuneration, 87 (48%) and 58 (32%) indicated they would be more likely to review a manuscript if they were offered £50 or a 12-month online subscription to one of our journals, respectively. However, approximately one-third said these incentives would not make them more likely to review (60 [33%] payment; 71 [39%] journal subscription), and 30 (16%) and 48 (26%), respectively, were unsure. Half of the participants (n = 93 [51%]) thought offering the choice of these incentives would help attract a more diverse sample of reviewers, 69 (38%) were unsure, and 15 (8%) did not. Views on remuneration were divergent (**Table 25-0908**) with some seeing it as a recognition of value and others considering it unnecessary. While 107 (58%) had no concerns about introducing payment for patient and public reviewers, 37 (20%) did have concerns and 33 (18%) were unsure. Concerns included potential conflicts of interest, tax implications and effort of declaring, impact on means-tested benefits, effect on review quality and integrity, people becoming reviewers for the wrong reasons, and high foreign exchange banking fees. Others viewed the remuneration as a modest token unlikely to influence quality. Respondents emphasized the importance of optional incentives to accommodate individual preferences and suggested monitoring the quality of impact after implementation of remuneration.

Conclusions *BMJ*'s patient and public reviewers hold diverse views on reviewer remuneration, highlighting the importance of providing flexible, optional incentive choices to accommodate varying individual needs, values, and preferences.

Table 25-0908. Examples of Comments About Remuneration for Review and its Impact From Patient and Public Reviewers at *BMJ*

Payment
"I worked in PPI for NIHR for some years and we considered payment for reviews an appropriate step, in the same way as we rewarded public funding committee contributors."
"I believe strongly in the principle of compensating patients for their time both as a measure of the worth of their insight and as a means of reducing barriers to their participation in such activities."
"Most patients are in financial precarity and offering payment might mean they actually can participate."
"It is important that those from the seldom heard community have the opportunity to contribute and often it is the lack of reimbursement that prevents them from doing so."
"It's nice to be acknowledged and of course everyone likes receiving money. But I can't understand why you're doing this—it's not necessary."
"While I am happy to receive a reward, I do wonder if it is necessary. I appreciate the opportunity to review *BMJ* publications, and personally, no reward is needed."
Journal subscription
"I very much enjoy access to other health journals as I feel it helps me improve as reviewer and in my PPI roles."
"Access to journals more valuable to me than £50."
Concerns about remuneration
"Potential for the journal to inadvertently exploit people, eg, people who have very low income may accept reviews when they aren't well enough because they need the money."
"Although the £50 compensation is within the allowed amount for most benefits, the process of legally receiving the money can sadly mean hours of online administration and possibly stopping of benefit payment temporarily."
"Having to involve HMRC is a definite disincentive to me."
"It may affect the integrity and quality of reviews."
"There is a risk of attracting reviewers who are more interested in reward than in contributing to scientific progress."
Need for continuous evaluation
"It is very important to continually evaluate and review whatever reward system you adopt."
"Perhaps good to try out this reward for a certain time and evaluate if it indeed meets the anticipated goals the *BMJ* has defined."

Abbreviations: HMRC: HM Revenue & Customs; NIHR, National Institute for Health and Care Research; PPI, patient and public involvement.

References

1. Asante K. Remuneration of African patient partners is an important tool for health justice. *BMJ*. 2024;387:q2675. doi:10.1136/bmj.q2675

2. Schroter S, Price A, Flemyng E, et al. Perspectives on involvement in the peer-review process: surveys of patient and public reviewers at two journals. *BMJ Open*. 2018;8(9):e023357. doi:10.1136/bmjopen-2018-023357

3. Doble E, Schroter S, Price A, et al. *The BMJ* will remunerate patient and public reviewers. *BMJ*. 2024;387:q2581. doi:10.1136/bmj.q2581

[1]*BMJ,* BMA House, Tavistock Square, London, UK, sschroter @bmj.com; [2]Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK; [3]Patient and public reviewer, *BMJ,* London, UK; [4]Dartmouth Health, NH, US; [5]Colorado School of Public Health, CO, US.

Conflict of Interest Disclosures Sara Schroter and Sophie Cook are employed by *BMJ*, Emma Doble is a freelance patient editor for *BMJ*, Amy Price was a freelance patient editor for *BMJ*, and Rebecca Harmston is a patient and public reviewer for *BMJ*.

Acknowledgment We thank Layla Abdulbaki, Shahad Al Mashjari, Shahd Rihawi, Alaa Almuhaini, Hygieia Jacob, Munya Alshaalani, Aliaa Bhahzad, Manvi Upadhaya, Zainab Almutawa, Sarah Khan, Adriana Goracci, and Adrian Yee for piloting the questionnaire, providing feedback on the clarity of the questions, and reviewing the work for readability and relevance. We appreciate all of *BMJ*'s patient and public reviewers who participated in the study.

Friday, September 5

Use of AI to Assess Quality and Reporting

Natural Language Processing to Assess the Role of Preprints in COVID-19 Policy Guidance

Nicholas G. Evans,[1] Samuel Angelli-Nichols,[1] Emma Chang-Rabley,[2] Yara Omar,[3] Rachel Nas,[4] Mikaela Finnegan,[4] Rocco Casagrande,[4] Emily E. Ricotta[5]

Objective To understand the role of preprint copies of scientific papers in public health policy guidance during the COVID-19 pandemic and the potential effects of changes in published versions of record (postprints) on that guidance.[1,2]

Design We extracted preprint citations from the Department of Homeland Security Master Question List from February 12, 2020, to February 9, 2021, and the National Institute for Occupational Safety and Health weekly COVID-19 Report from November 6, 2020, to September 17, 2021. Text of the abstract from each preprint was parsed and compared with its postprint using a neural conditional random field model for sentence alignment.[3] Identified sentence-level differences were subject to review and adjudication by the research team. Substantive changes were categorized as either nonsignificant numerical changes, contextual updates, or significant changes. Significant changes were compared against the text of the policy guidance to determine whether the postprint result would have resulted in changed guidance and whether guidance was updated after publication.

Results Our algorithm had a sensitivity of approximately 100% in detecting significant sentence-level changes when compared against a human adjudicator. Of 600 preprints extracted, 486 (81.0%) had associated postprints published during the study period. Our sample had a higher publication rate than global estimates of COVID-19 preprints (21.1%; odds ratio, 3.53; 95% CI, 3.10-4.02). Over time, guidance was slightly more likely to incorporate new postprints than preprints (odds ratio, 1.02; 95% CI, 1.01-1.03). Our model flagged 10,483 sentence-level changes across 464 preprints for review. Of these 464 papers, 329 (70.9%) required adjudication for the significance of their changes and 105 (22.6%) contained changes our team found could potentially change policy guidance. Significant changes included, among others, conclusions about the effectiveness of clinical interventions or nonpharmaceutical interventions to slow the spread of COVID-19, basic epidemiological properties, and the relative efficacy of vaccine candidates against variants of concern. Of 105 significant papers, 44 (41.9%) impacted policy guidance. Guidance that was impacted by significant changes tended to concern therapeutics and public health measures and, less often, changes resulting from the emergence of variants. The remainder involved guidance that lacked granularity of public health guidance relative to the kind of change in the paper's history. This was followed by a tendency toward conservative public health guidance when presented with competing evidence sources, which meant a single paper's change did not alter a recommendation. Only 9 papers containing policy-relevant changes were updated.

Conclusions Preprints can provide critical early information to policymakers but can generate unclear or misleading advice in key areas that may persist if not subject to clear and timely updates. While preprints in our sample tended to be of publishable quality, more work is needed to establish how to ensure this quality in future outbreak response.

References

1. Nelson L, Ye H, Schwenn A, Lee S, Arabi S, Hutchins BI. Robustness of evidence reported in preprints during peer review. *Lancet Global Health*. 2022;10(11):e1684-e1687. doi:10.1016/S2214-109X(22)00368-0

2. Collins, Alexander R. Reproducibility of COVID-19 pre-prints. *Scientometrics*. 2022;127(8):4655-4673. doi:10.1007/s11192-022-04418-2

3. Jiang, Maddela M, Lan W, Zhong Y, Xu W. Neural CRF model for sentence alignment in text simplification. *arXiv*. Preprint posted online ay 5, 2020. doi:10.48550/arXiv.2005.02324

[1]University of Massachusetts Lowell, Lowell, MA, US, nicholas_evans@uml.edu; [2]Emory University, Atlanta, GA, US; [3]Boston University, Boston, Massachusetts, US; [4]Deloitte, Rosslyn, VA, US; [5]Uniformed Services University of the Health Sciences, Bethesda, MD, US.

Conflict of Interest Disclosures None reported.

Funding/Support The initial funding of this work was provided through the Intramural Division of the National Institute for Allergy and Infectious Diseases. Nicholas G. Evans received funding from the Greenwall Foundation and from the US Air Force Office of Scientific Research, which provided time for the conduct of this research.

Role of Funder/Sponsor The funder had no role in the design, conduct, or reporting of this work.

Disclaimer Emily E. Ricotta is an employee of the US Department of Defense. This work does not represent the views of her organization or of the US government.

Additional Information Emily E. Ricotta is a co–corresponding author (emily.ricotta@usuhs.edu).

Leveraging Large Language Models for Assessing the Adherence of Randomized Controlled Trial Publications to Reporting Guidelines

Lan Jiang,[1] Xiangji Ying,[2] Mengfei Lan,[1] Andrew W. Brown,[3] Colby J. Vorland,[4] Evan Mayo-Wilson,[2] Halil Kilicoglu[1]

Objective SPIRIT[1] and CONSORT[2] guidelines recommend the minimum information to report in randomized clinical trial (RCT) protocols and results reports. Natural language processing (NLP) offers a promising approach for assessing whether manuscripts include the recommended information, the method used, and what was found. We annotated RCTs for adherence to SPIRIT 2013 and CONSORT 2010 and used large language models (LLMs) for automated assessment.

Design We annotated 100 protocol/results pairs (200 articles). We included parallel RCTs registered on ClinicalTrials.gov and published in PubMed Central[3] from January 2011 to August 2022. For inclusion criteria and search strategy, see https://osf.io/nefa9. Annotators rated 119 questions (85 check 1, 34 check all that apply), each related to a specific SPIRIT (n = 95) or CONSORT (n = 71) item or subitem. Questions assessed whether recommended information was reported and what specifically the trial did or found. For example, "Does the manuscript report any results for non-systematically assessed harms?" (Yes, No, or Cannot tell) was check 1; "If the study investigates a drug intervention or the intervention contains a drug component, does the manuscript mention the dosing schedule of the drug?" (Dose, Frequency, How long the drug will be taken, None reported, Not applicable, or No text to assess) was check all that apply. Two experts independently rated and adjudicated the protocols and results articles for 25 trials; the rest were rated by 1 expert. Protocol/results pairs were split 70:10:20 into training, validation, and test sets, respectively. To assess what the RCTs reported, we prompted GPT-4o with each question and its corresponding response options to elicit responses for every question for each article. We assessed model performance on the test set using F1 score, the harmonic mean of positive predictive value, and sensitivity (range: 0-1, higher is better). The mean F1 score measured overall performance across all questions. We compared results to a baseline that selected the most common response for each question in the training set.

Results There was no information for some questions in most articles. For example, only 4% reported informed consent materials. Even commonly reported items often lacked critical details (eg, 36% did not specify procedures for monitoring participant adherence to intervention). Interannotator agreement was 0.94 (Cohen κ). F1 score comparison for the GPT-4o model and the baseline is shown in **Table 25-1019** (0.856 vs 0.764). The scores ranged from 0.495 to 1 for GPT-4o and from 0.287 to 1 for baseline. Our approach performed comparably on questions related to both the SPIRIT[1] and CONSORT[2] guidelines, with slightly better results for SPIRIT.

Table 25-1019. Results by Question Type and Guideline Category With 95% Confidence Intervals

	Mean (95% CI)	
	Baseline	GPT-4o
Check 1	0.822 (0.789-0.853)	0.862 (0.834-0.889)
Check all that apply	0.620 (0.549-0.689)	0.840 (0.794-0.882)
SPIRIT	0.795 (0.757-0.831)	0.865 (0.838-0.891)
CONSORT	0.739 (0.691-0.785)	0.850 (0.819-0.880)

Conclusions LLMs hold promise for evaluating RCT adherence to reporting standards. Our questions might help pinpoint missing items and identify RCT characteristics. LLM-based tools could support peer review workflows, and further user engagement could help identify the SPIRIT and CONSORT items users consider highest priority for systematic assessment.

References

1. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med*. 2013;158(3):200-207. doi:10.7326/0003-4819-158-3-201302050-00583

2. Schulz KF, Altman DG, Moher D; CONSORT Group. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *BMJ*. 2010;340:c332. doi:10.1136/bmj.c332

3. Kilicoglu H, Rosemblat G, Hoang L, et al. Toward assessing clinical trial publications for reporting transparency. *J Biomed Inform*. 2021;116:103717. doi:10.1016/j.jbi.2021.103717

[1]School of Information Sciences, University of Illinois, Urbana-Champaign, Urbana-Champaign, IL, US, lanj3@illinois.edu; [2]Gillings School of Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, NC, US; [3]University of Arkansas for Medical Sciences and Arkansas Children's Research Institute, Little Rock, AK, US; [4]School of Public Health, Indiana University, Bloomington, IN, US.

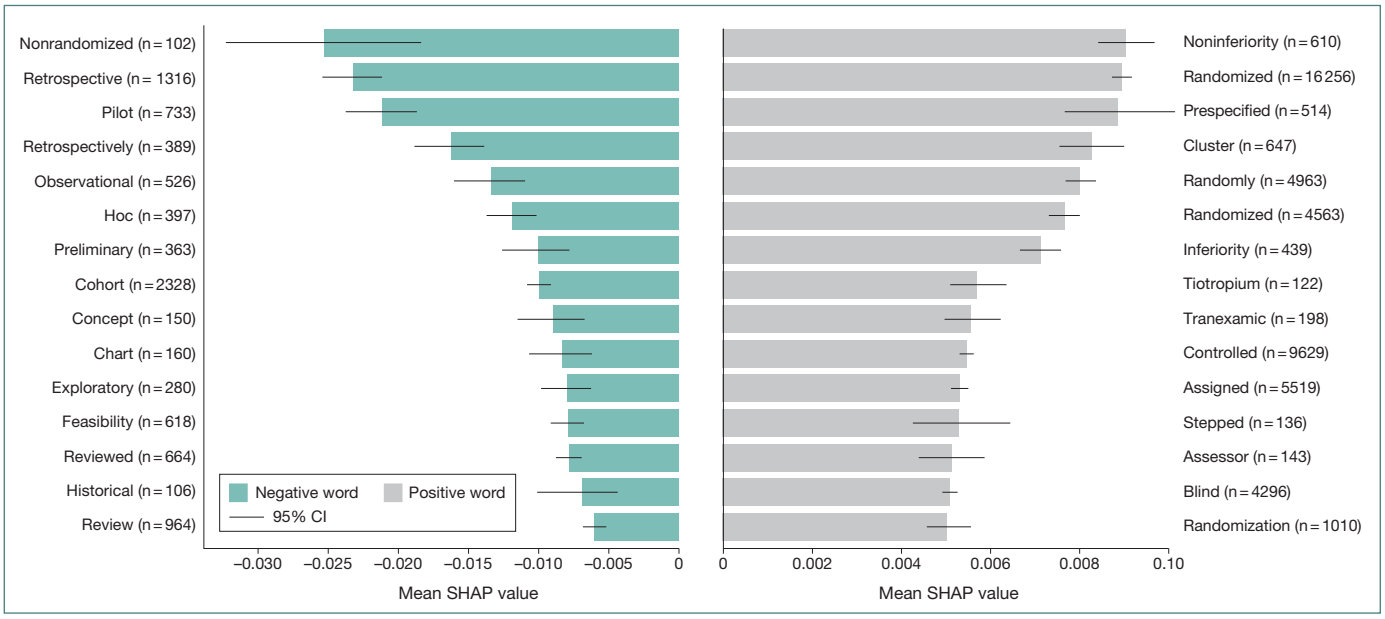
Conflict of Interest Disclosures None reported.

Funding/Support This study was supported by the US National Library of Medicine of the National Institutes of Health under award number R01LM014079.

Role of the Funder/Sponsor The funder had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Additional Information Halil Kilicoglu is a co–corresponding author (halil@illinois.edu). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Figure 25-0848. Mean SHAP Value of the 15 Words With the Largest Average Positive or Negative Contributions



The following 4 datasets were included: PLUS validate, PLUS test, PLUS 2024, and Clinical Hedges. Error bars represent the 95% CI. n indicates the number of words that occur across all samples.

Understanding How a Language Model Assesses the Quality of Randomized Controlled Trials: Applying Shapley Additive Explanations to Encoder Transformer Classification Models

Fangwen Zhou,[1] Muhammad Afzal,[2] Rick Parrish,[1] Ashirbani Saha,[3] R. Brian Haynes,[1] Alfonso Iorio,[1,4] Cynthia Lokker[1]

Objective Deep learning models for classifying published clinical literature to support critical appraisal have garnered wide attention.[1] To automate the critical appraisal workflow for McMaster's Premium Literature Service (PLUS), a gold-standard database manually labeled by experts was used to develop well-performing models to classify the rigor of randomized controlled trials (RCTs).[2] However, due to the complexity of language models, the lack of transparency remains an important concern. This study explored Shapley Additive Explanations (SHAP)[3] to better understand how a deep learning model makes classification decisions.

Design Details regarding the development and evaluation of the rigor classifiers are published elsewhere.[2] Briefly, classifiers were trained with titles and abstracts of original RCTs, that met or did not meet PLUS criteria for rigor (randomization, ≥10 participants per group, ≥80% follow-up, clinically important outcomes, and preplanned subgroup analyses, if applicable). Models were trained on 53,219 PLUS articles from 2003 to 2023, randomly split 80:10:10 into train, validate, and test sets. Articles in Clinical Hedges, a similar database that preceded PLUS, and PLUS articles from 2024 were used for external testing. A top-performing BioLinkBERT model, with an area under the receiver operating characteristic curve of 94%, was selected for this study. The SHAP partition explainer determined important tokens (words/subwords) for articles from the validate and test datasets, Clinical Hedges, and PLUS 2024. Tokens were combined into words for ease of interpretation, and their SHAP values were summed. The mean SHAP values, which indicate the aggregated average marginal contribution to model output over all samples, for the most impactful words with 100 or more occurrences were examined.

Results Overall, 6,207,935 words were analyzed, of which 49,387 were unique. **Figure 25-0848** shows the 15 most impactful unique words for positive (rigorous) and negative (nonrigorous) classes. Terms such as *noninferiority* positively influenced rigor predictions with a mean SHAP value of 0.00904 (95% CI, 0.00844-0.00965), whereas *nonrandomized* (mean SHAP value, −0.02510; 95% CI, −0.03204 to −0.01816) had negative impacts. These results generally align with the manual appraisal criteria. However, tokens apparently unrelated to the criteria, such as tiotropium (mean SHAP value, 0.00571; 95% CI, 0.00508-0.00634), also had sizable impacts, indicating that they may have been correlated with higher rigor during manual appraisal. These patterns were learned and applied by the model, indicating a certain degree of overfitting.

Conclusions This study demonstrates that SHAP helps understand which features influence a deep learning model's rigor assessment of an RCT. Identifying influential features increases confidence in the model assessments and generalizability and reduces unease about the black-box nature of these models. Future work should explore SHAP with other critical appraisal tools and datasets and potentially integrate it into machine learning systems to improve user trust and model accountability in evidence synthesis and critical appraisal workflows.

References

1. Lokker C, Bagheri E, Abdelkader W, et al. Deep learning to refine the identification of high-quality clinical research articles from the biomedical literature: performance

evaluation. *J Biomed Inform*. 2023;142:104384. doi:10.1016/j.jbi.2023.104384

2. Zhou F, Parrish R, Afzal M, et al. Benchmarking domain-specific pretrained language models to identify the best model for methodological rigor in clinical studies. *J Biomed Inform*. 2025;166:104825. doi:10.1016/j.jbi.2025.104825

3. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *arXiv*. Preprint posted online May 22, 2017. doi:10.48550/arXiv.1705.07874

[1]Health Information Research Unit, Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada, lokkerc@mcmaster.ca; [2]Faculty of Computing, Engineering and the Built Environment, Birmingham City University, Birmingham, UK; [3]Department of Oncology, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; [4]Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada.

Conflict of Interest Disclosures McMaster University, a nonprofit public academic institution, has contracts through the Health Information Research Unit under the supervision of Alfonso Iorio and R. Brian Haynes. These contracts involve professional and commercial publishers to provide newly published studies, which are critically appraised for research methodology and assessed for clinical relevance as part of the McMaster Premium Literature Service (McMaster PLUS). Cynthia Lokker and Rick Parrish receive partial compensation, and R. Brian Haynes is remunerated for supervisory responsibilities and royalties. Ashirbani Saha, Fangwen Zhou, and Muhammad Afzal are not affiliated with McMaster PLUS.

Funding/Support Fangwen Zhou was funded through the Mitacs Business Strategy Internship grant (IT42947) with matching funds from EBSCO Canada.

Role of the Funder/Sponsor The funders 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, and decision to submit the abstract for presentation.

Acknowledgment We thank the Digital Research Alliance of Canada for providing the computational resources.

Using GPT to Identify Changes in Clinical Trial Outcomes Registered on ClinicalTrials.gov

Xiangji Ying,[1] Colby J. Vorland,[2] Kiran Ninan,[1] Jean-Pierre Oberste,[1] Andrew W. Brown,[3] Riaz Qureshi,[4] Sirui Zhang,[5] Nicholas J. DeVito,[6] Matthew Page,[7] Ian J. Saldanha,[8] Halil Kilicoglu,[9] Evan Mayo-Wilson[1]

Objective Registries, such as ClinicalTrials.gov, and guidelines, including SPIRIT 2025 (Standard Protocol Items: Recommendations for Interventional Trials) and CONSORT 2025 (Consolidated Standards of Reporting Trials), define clinical trial outcomes using 5 elements: domain, measurement, metric, aggregation method, and time point. Changes between prospective registration and results reporting can introduce bias. Readers can manually compare trial documents, but doing so is resource intensive. Automated methods could facilitate checking and improve peer review. We used a chatbot to define outcomes registered on ClinicalTrials.gov, identify changes between prospective registrations and registry results, and describe those changes.

Design We conducted a cross-sectional study using prospectively registered, completed randomized clinical trials from ClinicalTrials.gov (January 2000-January 2024). Building on the 5-element framework, we developed rules and categories to define outcomes and outcome changes. We used GPT-4o, o1, and o3 mini (Open AI) and developed and optimized prompts using a training set of 225 trials (2221 outcomes). We validated performance on 150 trials (1459 outcomes). We divided the task into 16 structured subtasks in the following steps: (1) defining outcomes in both versions, (2) matching outcomes, and (3) detecting changes in outcome elements. We provided cosine and Jaccard distances to help the chatbot match and compare outcomes. We evaluated all outcomes using the 3 chatbot models in January 2025. We selected the answer given by at least 2 of the models. Two human raters independently evaluated the chatbot results. We present preliminary findings based on unreconciled ratings of changes (final findings will be presented at the conference).

Results The 150 validation trials reported a mean (SD) of 6.9 (5.0) outcomes in prospective registrations and 8.4 (6.0) outcomes in final versions. The final versions included 832 outcomes matched with prospective registrations and 428 additional outcomes and omitted 199 outcomes from prospective versions. We achieved 99.8% (95% CI, 99.4%-99.9%) accuracy in matching outcomes between versions (**Table 25-1142**). Accuracy for detecting changes was 87.9% (95% CI, 85.5%-90.1%) for measurement, 88.1% (95% CI, 85.7%-90.2%) for metric, 87.2% (95% CI, 84.8%-89.4%) for aggregation method, 94.5% (95% CI, 92.7%-95.9%) for cutoff, and 94.8% (95% CI, 93.1%-96.2%) for time point. On average, the language learning models completed the entire process (ie, matching, defining, comparing) in approximately 2 minutes per trial. It took humans 27 minutes per trial to evaluate the chatbot's response.

Conclusions A prompt-based approach was highly accurate in defining clinical trial outcomes and identifying outcome changes in ClinicalTrials.gov. This approach could be expanded to identify changes between registrations and manuscripts. Although it did not achieve perfect accuracy, our prompt-based approach could help editors and peer reviewers detect likely discrepancies that warrant further review.

[1]Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC, US, evan.mayo-wilson@unc.edu; [2]Department of Epidemiology and Biostatistics, Indiana University School of Public Health–Bloomington, IN, US; [3]Department of Biostatistics, University of Arkansas for Medical Sciences; Arkansas Children's Research Institute, Little Rock, AR, US; [4]Department of Ophthalmology, School of Medicine; Department of Epidemiology, School of Public Health, University of Colorado Anschutz Medical Campus, Denver, CO, US; [5]Department of Epidemiology, School of Public Health, Brown University, Providence, RI, US; [6]The Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK; [7]Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University,

Table 25-1142. Preliminary Accuracy of the Chatbot in Defining Outcomes and Identifying Outcome Changes[a] in 150 RCTs Registered on ClinicalTrials.gov[b]

	% Agreement (95% CI)	κ	PABAK[c]
Matching outcomes (1459 outcomes or outcome pairs)[d]			
Matched, added, or missing outcomes	99.8 (99.4-99.9)	0.99	0.99
Comparing elements (prospective vs final registration, 829 outcome pairs)			
Measurement	87.9 (85.5-90.1)[e]	0.62	0.76
Metric	88.1 (85.7-90.2)[e]	0.67	0.76
Aggregation method	87.2 (84.8-89.4)[e]	0.71	0.74
Cutoff	94.5 (92.7-95.9)[e]	0.60	0.89
Timepoint	94.8 (93.1-96.2)	0.90	0.90
Defining elements (prospective registration, 538 outcomes)			
Measurement	86.3 (83.0-89.0)	NE	0.72
Metric	90.9 (88.1-93.2)	0.87	0.82
Aggregation method	89.4 (86.5-91.9)	0.71	0.79
Cutoff	95.0 (92.8-96.7)	NE	0.90
Defining elements (final registration, 607 outcomes)			
Measurement	86.5 (83.5-89.1)	NE	0.73
Metric	91.6 (89.1-93.7)	0.87	0.83
Aggregation method	93.4 (91.1-95.3)[e]	0.87	0.87
Cutoff	95.6 (93.6-97.1)	NE	0.91

Abbreviations: NE, not estimable due to an infinite number of possible answers; PABAK, Prevalence-adjusted and bias-adjusted κ.

[a]We defined outcome changes for each element as more complete, less complete, or change in definition comparing the element in the final registration to that in the prospective registration.

[b]We manually rated the comparison results of all 150 trials. However, only the outcome definitions of 75 trials in the prospective and final registrations were rated by human reviewers.

[c]PABAK may be preferable over unweighted κ because certain categories (eg, no change) occur more frequently than other categories.

[d]Matching was only done by GPT-4o.

[e]When all 3 models provided different answers, 1 of their responses was randomly selected as GPT's final answer. The 3 models provided different answers for 21 outcomes (2.5%) regarding measurement comparison, 19 outcomes (2.3%) regarding metric comparison, 4 outcomes (0.5%) regarding aggregation method comparison, 11 outcomes (1.3%) regarding cutoff comparison, and 2 outcomes (0.2%) in the final registration regarding the definition of the aggregation method.

Melbourne, Victoria, Australia; [8]Departments of Epidemiology (Primary) and Health Policy and Management (Joint), Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, US; [9]School of Information Sciences, University of Illinois at Urbana-Champaign, IL, US.

Conflict of Interest Disclosures None reported.

Funding/Support National Library of Medicine, National Institutes of Health (R01LM014079).

Role of Funder/Sponsor The funder played no roles 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. The views expressed do not necessarily reflect the views of the National Institutes of Health.

Additional Information Our protocol is publicly available and we will share prompts, code, and datasets on the Open Science Framework (https://osf.io/2tyh3/).

Open Science, Availability of Protocols, and Registration

Perceived Risks and Barriers to Open Research Practices in UK Higher Education

Lukas Hughes-Noehrer,[1] Noémie Aubert Bonn[1]

Objective Open and transparent research practices are essential for high-quality research. Despite growing efforts to promote these practices through new funding policies, a rethinking of evaluation criteria, and investment in support and infrastructure, widespread adoption across disciplines remains challenging. Comprehensive data and analysis of the perceived risks and barriers to implementing open research practices within UK higher education institutions (HEIs) is still lacking.

Design In 2025, we conducted an inductive thematic analysis of openly available data from the UK Reproducibility Network Open and Transparent Research Practices Survey,[1] which was conducted between December 2022 and April 2023. The survey addressed 14 topics related to open research practices (eg, research coproduction, open-source code and software, reproducibility, open access, FAIR principles [findability, accessibility, interoperability, and reusability], and conflicts of interests), with respondents answering questions on practice, support, and approaches to each practice. This analysis focuses on responses to the open-text questions specifically referring to risks or barriers to researchers or to their fields of study across all 14 topics. Following SRQR guidelines,[2] we ensured transparency in our analytic process. One researcher analyzed the data and developed initial codes, which have been reviewed by a second coder, before agreeing on collaboratively refined themes through iterative discussion. We used NVivo, version 14 (Lumivero) for analysis and data management. Reflexivity was maintained through regular team debriefs. Ethical approval was not required as the data were publicly available and anonymized.

Results The survey received 2567 responses from research-active staff at 15 UK HEIs. Of these, 511 (20%) responses were complete. Participants reported barriers and risks to the 14 open research topics (**Table 25-0863**), which we grouped into 3 overarching categories: personal, systemic and institutional, and ethical and quality. In the personal category, respondents highlighted lacking motivation to undertake open research practices, with some perceiving that such practices should not rely on researchers and others perceiving such practices as a mere tick-box exercise. Respondents also feared that open research practices would expose errors and make them vulnerable to criticism. In the systemic and institutional category, respondents described barriers that made them unable to practice open research, such as methodological or disciplinary barriers. They also highlighted missing infrastructure and a lack of resources. Others also mentioned that institutions did not support them adequately, did not provide appropriate training, and did not recognize or incentivize open research practices, making it

Table 25-0863. Surveyed Open Research Practices and Their Top 3 Identified Risks and Barriers (Ordered Top to Bottom)

Open research practices[a]	Associated top 3 risks and barriers
Research coproduction	Disciplinary barriers; ethical concerns; problematic in collaborations
Conduct of open research consistent with relevant legal, ethical, and regulatory constraints	Time and resources; ethical concerns; disciplinary barriers
Transparent qualitative data practices	Disciplinary barriers; ethical concerns; time and resources
Defining the data, code, or other evidence before the start of data collection and analysis	Time and resources; disciplinary barriers; lack of training
Preregistration of research protocols	Disciplinary barriers; time and resources; lack of training
Use of open-source software created by others	Issues with infrastructure; lack of training; time and resources
Creation of open-source software	Time and resources; lack of training; disciplinary barriers
Version control of research products	Lack of training; time and resources; issues with infrastructure
Computational reproducibility of data analysis	Disciplinary barriers; time and resources; lack of training
Preparing data, code, or other evidence according to the FAIR principles	Lack of training; time and resources; disciplinary barriers
Guidelines for recognizing the specific substantive contribution of everyone involved in research projects	Lack of incentives or recognition; lack of standards or guidance; theft of credit or ideas
Declaring conflicts of interests	Disciplinary barriers; lack of standards or guidance; ethical concerns/lack of training
Publication of preprints	Lack of incentives or recognition; quality and innovation concern; theft of credit or ideas
Ensuring publications are open access	Time and resources; lack of incentives or recognition; ethical concerns

Abbreviations: FAIR, findability, accessibility, interoperability, and reusability; OA, open access.
[a]For a further description of practices, consult the FORRT Glossary (https://forrt.org/glossary/).

difficult for them to engage in open research. On an ethical and quality level, respondents worried that open research practices introduced new ethical concerns, or even that they detract from high-quality research and innovation.

Conclusions These findings underscore critical challenges hindering the adoption of open research practices in UK HEIs. Addressing these barriers requires enhanced institutional support, targeted training, and robust incentives to empower researchers. Mitigating these challenges could significantly improve the quality and transparency of academic research.

References

1. Hughes-Noehrer L, Aubert Bonn N, De Maria M, et al. UK Reproducibility Network open and transparent research practices survey dataset. *Sci Data*. 2024;11:912. doi:10.1038/s41597-024-03786-z

2. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*. 2014;89(9):1245-1251. doi:10.1097/ACM.0000000000000388

[1]Department of Computer Science, The University of Manchester, Manchester, UK, lukas.noehrer@manchester.ac.uk.

Conflict of Interest Disclosures None reported.

Funding/Support This work was partially funded by the Research England Development Fund. Research England had no further involvement in the research life cycle.

Acknowledgment We want to thank all members of the United Kingdom Reproducibility Network who have contributed to the Open and Transparent Research Practices survey.

Use of an Open Science Checklist and Reproducibility of Findings: A Randomized Controlled Trial

Ayu Putu Madri Dewi,[1] Melissa L. Rethlefsen,[2] Sara Schroter,[3,4] Florian Naudet,[5,6] Nicholas J. DeVito,[7] Constant Vinatier,[5] Inge Stegeman,[8,9] Mariska Leeflang,[1] Gowri Gopalakrishna[1,10]

Objective *The BMJ* is one of the first journals to require authors to share analytic codes from all studies as well as data from all trials,[1] but whether this enhances reproducibility has not been tested. We will assess whether adding an open science checklist to the journal peer review process improves the reproducibility of scientific findings in submitted manuscripts to *The BMJ* and *BMJ Open*.

Design This study is registered on Open Science Framework[2] and follows the SPIRIT guidelines. Reproducibility was defined as obtaining the same primary result when rerunning analyses using the original data and code. We used a 2-arm parallel randomized controlled trial design, equal allocation, and random block randomization (block sizes of 4, 6, and 8; generated in R version 4.3.2 [The R Foundation]). We hypothesized that the intervention will increase reproducibility of primary outcomes from 50% in the control arm to 70% in the intervention arm (90% power; $\alpha = .05$). Our minimum sample size is 260 manuscripts (minimum 130 manuscripts per arm). Eligible manuscripts include quantitative observational or experimental designs and systematic reviews containing a meta-analysis. In the intervention arm, the research team completed an open science checklist assessing whether a set of predefined open science items[3] (**Table 25-1009**) were reported in the

Table 25-1009. Checklist of Open Science to Improve Reproducibility in Biomedical Science

Science checklist item	Description	Yes/no/not applicable	Viewer's notes—indicate what and where the information is available
Availability of study protocol	Is study protocol publicly available?		
Preregistration of study	Has the study been preregistered (eg, at the Open Science Framework)?		
Open materials	Are relevant materials (eg, surveys, experimental setups) available for other researchers to inspect or reuse?		
Open data	Are relevant data available for other researchers to replicate?		
Open code	Is relevant code available for other researchers to reuse?		
Software or tools	Is any software or tool used in the study openly available or is there a clear explanation of how to access it?		
Open access reporting	Have the authors requested open access?		
Code availability	Is any code used for analysis provided in an accessible, open-source repository (eg, GitHub, Zenodo)?		
Transparent reporting	Is the appropriate reporting guideline checklist used?		
Reporting citations to materials	Are citations or persistent identifiers provided for materials?		
Reporting citations to data	Are citations or persistent identifiers provided for data?		
Reporting citations to code	Are citations or persistent identifiers provided for code?		
Reporting the preprint(s)	Is a preprint available?		
Additional comment(s):			

manuscripts. The completed checklist was shared with the authors as part of the standard peer review process. Manuscripts in the control arm followed the usual peer review process. The primary outcome was the difference in reproducibility of the authors' reported primary outcomes between the 2 arms, as assessed by blinded outcome assessors. The secondary outcome was the proportion of manuscripts with openly shared data and code. If the data, code, or methods were unavailable or only available on request, authors were not contacted, as they were blinded to the study, and the outcome assessors did not proceed. Randomization began on March 3, 2025, with outcome assessments conducted after the first manuscript revision.

Results By April 18, 2025, we had completed randomization of 403 manuscripts. This includes open science intervention in 22 of 43 manuscripts from *The BMJ* and 179 of 360 manuscripts from *BMJ Open*. As of June 5, 2025, 78 manuscripts had received a decision letter, including 39 rejections, while the remainder proceeded to reproducibility assessment. Analysis is ongoing.

Conclusions This study will provide robust evidence on whether incorporating an open science checklist into the peer review process improves the reproducibility of scientific findings.

References

1. Loder E, Macdonald H, Bloom T, Abbasi K. Mandatory data and code sharing for research published by *The BMJ*. *BMJ*. 2024;384:q324. doi:10.1136/bmj.q324

2. Dewi APM, Rethlefsen ML, Schroter S, et al. Measuring the efficacy of an intervention to improve reproducibility of scientific manuscripts: study protocol for a randomized controlled trial. OSF. Accessed July 2, 2025. https://osf.io/b5g6y/

3. Cobey KD, Haustein S, Brehaut J, et al. Community consensus on core open science practices to monitor in biomedicine. *PLoS Biol*. 2023;21(1):e3001949. doi:10.1371/journal.pbio.3001949

[1]Epidemiology and Data Science Department, Amsterdam UMC, Amsterdam, the Netherlands; [2]Health Sciences Library and Informatics Center, University of New Mexico, Albuquerque, NM, US; [3]BMJ Publishing Group, London, UK; [4]Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK; [5]Univeristy of Rennes, CHU Rennes, INSERM, EHESP, Irset (Institut de recherche en santé, environnement et travail), UMR_S 1085, Rennes, France; [6]Institut Universitaire de France, Paris, France; [7]Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK, nicholas.devito@phc.ox.ac.uk; [8]Department of Otorhinolaryngology and Head & Neck Surgery, University Medical Center Utrecht, Utrecht, the Netherlands; [9]Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands; [10]Department of Epidemiology, Faculty of Health, Medicine, and Life Sciences Maastricht University, Maastricht, the Netherlands.

Conflict of Interest Disclosures None reported.

Nonregistration, Discontinuation, and Nonpublication of Randomized Trials in Switzerland, the UK, Germany, and Canada: An Updated Meta-Research Study

Benjamin Speich,[1] Ala Taji Heravi,[1] Johannes M. Schwenke,[1] Christof M. Schönenberger,[1] Lena Hausheer,[1] Dmitry Gryaznov,[1] Jason W. Busse,[2,3] Manuela Covino,[1] Szimonetta Lohner,[4] Malena Chiaborelli,[1] Ruben Ramirez,[1] Ramon Saccilotto,[5] Erik von Elm,[6] Arnav Agarwal,[3] Julian Hirt,[7] David Mall,[1] Alain Amstutz,[1] Selina Epp,[1] Dominik Mertz,[3]

Anette Blümle,[8] Belinda von Niederhäusern,[9] Ayodele Odutayo,[10] Alexandra N. Griessbach,[1] Sally Hopewell,[11] Matthias Briel,[1,3] for the ASPIRE Study Group

Objective Previous studies found that approximately one-third of randomized clinical trials (RCTs) were discontinued prematurely and that the most common reason for stopping early was poor recruitment of participants. To minimize research waste, it is crucial that all RCTs are registered and make their results available. Hence, we aimed to (1) assess the fate of RCTs approved by ethics committees in 2016 in terms of nonregistration, discontinuation, and nonpublication and (2) examine RCT characteristics associated with discontinuation due to poor recruitment and nonpublication of RCT results.

Design We had access to 347 RCT protocols that were approved in 2016 by research ethics committees in the UK, Switzerland, Germany, and Canada. Key trial characteristics were extracted from approved trial protocols. Trial registration was verified using registration numbers from the protocol and by searching the World Health Organization International Clinical Trials Registry Platform, ClinicalTrials.gov, European Union Clinical Trials Register, ISRCTN, and Google; trials were deemed nonregistered if not found. We searched for full-text publications in PubMed, Google Scholar, and Scopus. Searches for both registration and publication used (1) full titles, (2) short titles, (3) study acronyms, and (4) the study population and intervention. All searches were conducted in duplicate (last search, July 2024). In case the status of an RCT was unclear, we contacted the corresponding ethics committee or the principal investigator. We reported the proportion of nonregistered RCTs, discontinued RCTs (including reason for early discontinuation), and nonpublished RCTs (considering peer-reviewed publications and results in trial registries).

Results Of the 347 included RCTs from 2016, 20 (5.8%) were nonregistered (industry-sponsored RCTs, 5 of 181 [2.8%]; nonindustry RCTs, 15 of 166 [9.0%]). Approximately one-third of RCTs (108 of 347 [31.1%]) were discontinued, most often due to poor recruitment (49 of 108 [45.4%]). A total of 276 of 347 RCTs (79.5%) made their results available at any source, 249 (71.8%) as a peer-reviewed publication, and 170 (49.0%) in the trial registry. Discontinued RCTs had lower result availability than completed RCTs (discontinued, 74 of 108 [68.5%]; completed, 202 of 226 [89.4%]). Results from industry-sponsored trials were more often available compared with nonindustry trials (92% vs 66%). This difference was driven by the fact that only 17 of 166 nonindustry RCTs (10.2%) reported results in trial registries compared with 153 of 181 industry trials (84.5%). The status of 4% of RCTs remained unclear despite our efforts to contact investigators.

Conclusions Nonregistration, premature discontinuation due to poor recruitment, and nonpublication of RCT results remain major challenges, especially for nonindustry trials.

[1]Division of Clinical Epidemiology, Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland, johannesmanteo.schwenke@usb.ch; [2]Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada; [3]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada; [4]MTA–PTE Lendület Momentum Evidence in Medicine Research Group, Medical School, University of Pécs, Pécs, Hungary; [5]Clinical Trial Unit, Department of Clinical Research, University of Basel and University Hospital Basel, Basel, Switzerland; [6]Cochrane Switzerland, Centre for Primary Care and Public Health (Unisante), University of Lausanne, Lausanne, Switzerland; [7]Research Center for Clinical Neuroimmunology and Neuroscience Basel, University Hospital Basel and University of Basel, Basel, Switzerland; [8]Clinical Trials Unit, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany; [9]Roche Pharma AG, Grenzach-Wyhlen, Germany; [10]University Health Network, Division of Nephrology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada; [11]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.

Conflict of Interest Disclosures Benjamin Speich and Matthias Briel received unrestricted grants from Moderna for the conduct of the COVERALL-2 and COVERALL-3 study. Benjamin Speich has received honoraria from Moderna and Roche for presenting study results not related to this work. Johannes M. Schwenke is paid by the Swiss National Science Foundation for work unrelated to this project. No other disclosures were reported.

Funding/Support The study is supported by the Swiss Federal Office of Public Health (Matthias Briel).

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.

Group Information The ASPIRE Study Group included Dmitry Gryaznov, Belinda von Niederhäusern, Benjamin Speich, Benjamin Kasenda, Elena Ojeda-Ruiz, Anette Blümle, Stefan Schandelmaier, Dominik Mertz, Ayodele Odutayo, Yuki Tomonaga, Alain Amstutz, Christiane Pauli-Magnus, Viktoria Gloy, Szimonetta Lohner, Karin Bischoff, Katharina Wollmann, Laura Rehner, Joerg J. Meerpohl, Alain Nordmann, Katharina Klatte, Nilabh Ghosh, Ala Taji Heravi, Jacqueline Wong, Ngai Chow, Patrick Jiho Hong, Kimberly McCord, Sirintip Sricharoenchai, Jason W. Busse, Lukas Kübler, Pooja Gandhi, Zsuzsanna Kontar, Julia Hüllstrung, Mona Elafy, Arnav Agarwal, Ramon Saccilotto, Alexandra N. Griessbach, Christof Schönenberger, Matthias Schwenkglenks, Giusi Moffa, Lars G. Hemkens, Sally Hopewell, Erik von Elm, and Matthias Briel.

Additional Information Matthias Briel is a co–corresponding author (matthias.briel@usb.ch).

Factors Associated With Improper Clinical Trial Registration, Registration Deficiencies, and Publication Status of Submissions to *The BMJ*

David Blanco,[1] Elizabeth Loder,[2,3] Sophie Cook,[3] Martí Casals,[4,5] Jordi Cortés,[6] Aïda Cadellans-Arroniz,[1] Victor Zárate,[1] Ella Hubbard,[3] Sara Schroter[3,7]

Objective In 2005, the International Committee of Medical Journal Editors (ICMJE) mandated that all clinical trials be properly registered—that is, registered in an approved registry at or before the enrollment of the first participant.[1] Despite improved registration rates, many trials published in journals claiming adherence to ICMJE registration policy

remain improperly registered.[2] This study aimed to (1) identify variables associated with improper registration; (2) examine types of registration deficiencies, publication status, and disclosure of registration issues in improperly registered trials; and (3) assess authors' claims of proper registration.

Design This observational study analyzed 239 improperly and 239 properly registered trials submitted to *The BMJ* (2019-2023). For objective 1, we collected data on trial design, setting, sample size, intervention type, number of authors, corresponding author region, mention of Consolidated Standards of Reporting Trials (CONSORT) guidelines, submission year, and funding source. For objective 2, we examined improperly registered trials submitted between 2019 and 2021, focusing on registration deficiency type (retrospective registration in approved registry, registration in unapproved registry, or no registration), registration delay (for retrospectively registered trials), and publication status. For published trials, we recorded Journal Impact Factor, ICMJE affiliation, time to publication, and disclosure of registration issues. For objective 3, we assessed authors' claims of prospective registration for improperly registered trials. Analyses included multivariable logistic regression and descriptive statistics.

Results Several variables were associated with reduced odds of improper registration: larger sample size (101-500 vs 1-100; odds ratio [OR], 0.43 [95% CI, 0.22-0.84]), corresponding authors from Oceania (reference: Europe; OR, 0.35 [95% CI, 0.14-0.82]), more authors (10 vs 1; OR, 0.53 [95% CI, 0.32-0.87]), CONSORT mention (OR, 0.22 [95% CI, 0.06-0.67]), recent submissions (2021-2023 vs 2019-2020; OR, 0.63 [95% CI, 0.42-0.96]), and funding (eg, nonprofit vs no funding; OR, 0.20 [95% CI, 0.09-0.41]). Trials with authors from Asia (OR, 1.75 [95% CI, 1.07-2.89]) had higher odds of improper registration. Of 176 improperly registered trials, 82.4% (n = 145) were retrospectively registered in approved registries (median delay, 6.54 months; Q1, 2.50 months; Q3, 18.67 months), 13.1% (n = 23) were unregistered, and 4.5% (n = 8) used unapproved registries. Most (88.1% [n = 155 trials]) were later published, including 1 in *The BMJ*. Among these, 89.0% (n = 138) appeared in journals with an Impact Factor (median, 5.39; Q1, 3.98; Q3, 10.40) and 62.0% (n = 96) in journals claiming adherence with ICMJE registration policy. Median time to publication was 12 months (Q1, 8 months; Q3, 19 months). Only about one-sixth explicitly acknowledged the registration issue at the time of publication. Of 72 responses to *The BMJ*'s submission question on prospective registration (2021-2023), 83.3% (n = 60) incorrectly claimed compliance.

Conclusions We identified variables associated with improper trial registration. Retrospective registration was common, and most improperly registered trials rejected by *The BMJ* were later published in other ICMJE-affiliated journals. ICMJE journals should strengthen processes to better identify and reject improperly registered trials.

References

1. International Committee of Medical Journal Editors. Clinical trials. Accessed January 23, 2025. https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html

2. 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

[1]Department of Physiotherapy, Universitat Internacional de Catalunya, Barcelona, Spain, dblanco@uic.es; [2]Department of Neurology, Brigham and Women's Hospital, Boston, MA, US; [3]*The BMJ*, London, UK; [4]National Institute of Physical Education of Catalonia, University of Barcelona, Barcelona, Spain; [5]Sport and Physical Activity Studies Centre, Faculty of Medicine, University of Vic - Central University of Catalonia, Barcelona, Spain; [6]Biostatistics and Bioinformatics Research Group, Department of Statistics and Operations Research and Institute for Research and Innovation in Health, Universitat Politècnica de Catalunya, Barcelona, Spain; [7]Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, London, UK.

Conflict of Interest Disclosures Sophie Cook and Sara Schroter are full-time employees of the BMJ Group and Elizabeth Loder is a part-time employee.

Registered Clinical Trial Trends in East Asia and the United States, 2014 to 2025

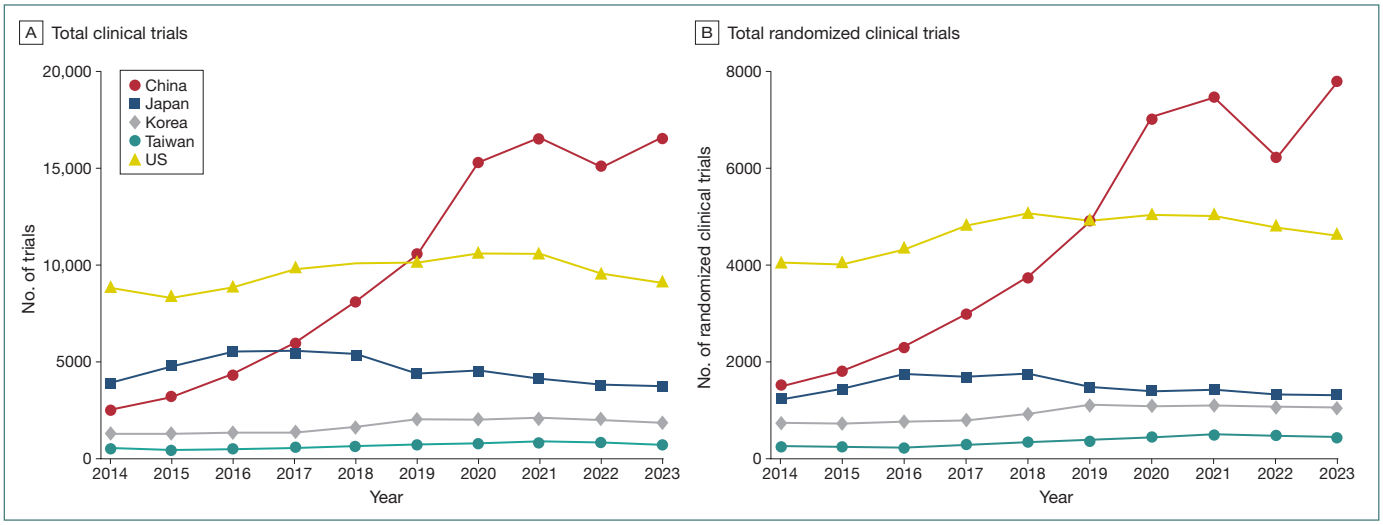
Eun Hye Lee,[1,2,3,4] San Lee,[1,3,5] Jae Il Shin,[3,4,6] John P. A. Ioannidis[1,7,8,9]

Objective East Asia has become a major hub for clinical trials.[1] Although a few studies have compared ClinicalTrials.gov with local registries in specific disease domains, comprehensive comparisons remain limited.[2] This study analyzes registered clinical trial trends in East Asia during 2014 to 2025 and compares them with trends in the US.

Design We extracted clinical trial data through the World Health Organization–operated International Clinical Trials Registry Platform (ICTRP), which integrates ClinicalTrials.gov and local registries. Because Taiwan's registry is not part of the ICTRP, we collected its data separately. We analyzed overall clinical trial trends and filtered randomized clinical trials (RCTs), categorizing them by location, target size, and disease category. Trials conducted in only 1 country were categorized as domestic, whereas those conducted in 2 or more countries were classified as international. We explored the proportion of studies registered prospectively, defined as within 1 month of the trial start date.[3]

Results China experienced rapid growth in clinical trials, surpassing both Japan and the US in total trials and RCTs (**Figure 25-1020**). In China, the number of registered trials increased markedly from 2578 total trials (1521 RCTs) in 2014 to 16,612 (7798 RCTs) in 2023. In contrast, the US showed a more modest increase from 8841 trials (4056 RCTs) in 2014 to 9100 (4619 RCTs) in 2023. Except for China, all other countries experienced a postpandemic decline, with annual trial registrations decreasing by approximately 3% to 10% in

Figure 25-1020. Change in Number of Clinical Trials



Total trials include trials from ClinicalTrials.gov and local registries for each country (the US includes only ClinicalTrials.gov data).

recent years. Neoplastic diseases and cardiovascular and metabolic diseases accounted for the largest percentages, each representing approximately 15% to 25% of all registered trials. In the US, mental health trials increased from 11.9% in 2014 to 16.0% in 2023, reflecting growing interest in this area. In 2023, China's RCTs were predominantly domestic, with international trials making up only 1.8% (140) of the total, while 13.1% (603) of RCTs in the US were international. Domestic RCTs were predominantly small-sized trials (<100 participants), while international RCTs were more commonly medium-sized (100-499 participants) or large-sized (≥500 participants) trials. In 2023, 75% of China's RCTs (5841 of 7798) and 81% of Japan's RCTs (1081 of 1339) were domestic trials registered in local registries, highlighting the predominant use of local registries in both countries. Although the trend has increased over the past decade, a considerable percentage of RCTs, ranging from 15% to 35% across countries, were still not registered prospectively in 2023. We recorded some inconsistencies and missing information across existing registries. Updated data to mid-2025 will be presented at the congress.

Conclusions The study highlights the increasing prominence of East Asia, particularly China, in registered clinical trials, though with a primarily domestic focus. Improvements in existing trial registries covering East Asia are desirable.

References

1. Ali S, Egunsola O, Babar ZUD, Hasan SS. Clinical trials in Asia: a World Health Organization database study. *Perspect Clin Res*. 2019;10(3):121-124. doi:10.4103/picr.PICR_109_18

2. Doi M, Yukawa K, Sato H. Characteristics of Asian 4 countries on cancer clinical trials registered in the international clinical trials registry platform between 2005 and 2018. *Chin Clin Oncol*. 2021;10(3):28. doi:10.21037/cco-21-17

3. Klatte K, Sluka c, Gloy V, et al. Towards full clinical trial registration and results publication: longitudinal meta-research study in Northwestern and Central Switzerland. *BMC Med Res Methodol*. 2023;23(1):27. doi:10.1186/s12874-023-01840-9

[1]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, US, jioannid@stanford.edu; [2]Division of Pulmonology, Allergy and Critical Care Medicine, Department of Internal Medicine, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin-si, Gyeonggi-do, South Korea; [3]The Center for Medical Education Training and Professional Development, Yonsei Donggok Medical Education Institute, Seoul, Korea; [4]Severance Underwood Meta-Research Center, Institute of Convergence Science, Yonsei University, Seoul, South Korea; [5]Department of Psychiatry and the Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, South Korea; [6]Department of Pediatrics, Yonsei University College of Medicine, Seoul, South Korea; [7]Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, US; [8]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, US; [9]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, US.

Conflict of Interest Disclosures John P. A. Ioannidis is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Funding/Support The work of John P. A. Ioannidis is supported by an unrestricted gift from Sue and Bob O'Donnell to Stanford University.

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

Additional Information A language model based on Bidirectional Encoder Representations from Transformers (BERT) was employed on December 30, 2024, to classify disease categories in RCTs. Specifically, PubMedBERT, a version pretrained on biomedical literature from PubMed, was used to automate the disease classification of clinical trials. San Lee takes responsibility for the integrity of the content. Eun Hye Lee and San Lee contributed equally to this study.

Open Science and Data Sharing

Researcher Adherence to Journal Data Sharing Policies: A Meta-Research Study

Aidan Christopher Tan,[1,2] Yiyi Lin,[1] Michellie Lian,[1] Zhilin Ren,[1] Tony Lian,[1] Vincent Yuan,[3] Angela Webster,[1] Anna Lene Seidler[1]

Objective We aimed to describe researcher adherence to journal data sharing policies across health research. This included describing what factors increase intention to share data and how journal policies requiring data sharing impact intended data sharing.

Design This was a cross-sectional study of all original research published in the highest-impact medical journals in 2022. Journals were included if they ranked among the top 5 in Impact Factor in 59 fields of medicine in the 2020 Journal Citation Reports, had a data sharing policy that either required or recommended sharing data, and published original research. Journal websites were searched to identify all original research published in 2022. Data were manually collected on study characteristics (study type, industry involvement, and COVID-19 relevance) and data sharing plans (data sharing statements and data sharing intentions extracted from publicly available research protocols and articles). Articles were descriptively analyzed by these journal data sharing policies, study characteristics, and data sharing plans.

Results Of 134 journals included, 36 (27%) required data sharing and 98 (73%) recommended data sharing. The analysis included 1868 interventional studies (1383 randomized clinical trials [74%] and 485 nonrandomized trials [26%]) and 10,368 observational studies (4814 cohort studies [46%], 4166 cross-sectional studies [40%], 1100 case-control studies [11%], and 288 case series/reports [3%]). Publicly available research protocols were available for 1993 of 10,243 studies (16%), of which 1153 (58%) had an initial data sharing statement. Only 1023 interventional studies (55%) and 4511 observational studies (44%) in journals that recommended or required data sharing actually intended to share data. Most studies only intended to share data underlying the published results with researchers, for purposes and by mechanisms at the discretion of and subject to approval by the principal investigator and without supporting documents or specified timeframes. Factors associated with increased intention to share data included journal policies that required data sharing and data sharing statements as well as industry involvement and COVID-19 relevance for interventional studies. For journals with policies that required data sharing, researchers were more likely to share more data (ie, all data collected during the study) and supporting documents (ie, supporting documents) in a timelier manner (ie, immediately following publication and indefinitely) to more people (ie, to anyone who wishes to access the data) for more purposes (ie, for any purpose) and by easier means (ie, with unrestricted access through a third-party website).

Conclusions Journals that required data sharing had higher rates of studies that intended to share data compared with journals that only recommended data sharing. However, most studies that intended to share data had restrictions on sharing. In addition to requiring data sharing and data sharing statements, journals should define and explain data sharing and individual participant data, better review data sharing statements and the reasonableness of justifications to not share data, and provide additional guidance on the operationalization of data sharing in accordance with best practice.

[1]NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia, aidan.tan@sydney.edu.au; [2]Sydney School of Public Health, University of Sydney, Sydney, Australia, [3]University of New South Wales, Sydney, Australia.

Conflict of Interest Disclosures None reported.

Funding/Support This work was supported by a research contract from CT:QI.

Role of the Funder/Sponsor CT:QI had no role in the study design, conduct, analysis, or reporting.

A Funder-Led Intervention to Increase the Sharing of Data, Code, Protocols, and Key Laboratory Materials

Robert Thibault,[1,2] Dana E. Cobb-Lewis,[1,2] Matt Lewis,[1,2] Devin Snyder,[1,2] Cornelis Blauwendraat,[1,2] Sonya Dumanis[1,2]

Objective We examined whether a funder-led open science policy and compliance workflow could increase the deposition and unambiguous identification of research data, code, protocols, and key laboratory materials (eg, cell lines, antibodies).

Design A cross-sectional study was conducted on articles supported by the Aligning Science Across Parkinson's (ASAP) Collaborative Research Network (CRN) and published between January 1, 2024, and April 30, 2025 (N = 102). The study was exploratory, and thus was not registered, and does not present inferential statistics. ASAP has a thorough Open Science Policy[1] and Compliance Workflow,[2] which require grantees to send manuscript drafts to ASAP staff, post preprints, deposit research outputs, and unambiguously identify research inputs. The workflow is managed by an ASAP staff member who integrates automated and manual assessments to provide a grantee with systematic feedback outlining the actions required to align their manuscript draft with the ASAP Open Science Policy.

Results Between the version of a manuscript first shared with ASAP staff and the associated final publication, there were substantial increases in the deposition of newly generated datasets (21% for the first version to 88% for the published version), unambiguous identification of reused datasets (72% to 86%), deposition of newly generated code (15% to 72%), unambiguous identification of software used (35% to 79%), registration of newly generated key lab materials (22% to 78%), unambiguous identification of key lab materials used (44% to 87%), and deposition of newly

generated protocols and unambiguous identification of existing protocols (35% to 81%; data on protocols was collapsed) (**Table 25-0892**). Of the included publications, 98% (100 of 102) had an associated preprint. Preprints were posted a median (IQR) of 4 (−7 to 50) days before submission to the journal in which they were eventually published and 236 (150-328) days before they were published in a journal. To facilitate this workflow, grantees share each manuscript alongside a Key Resource Table that lists persistent identifiers for the associated datasets, code, software, protocols, and key lab materials. Collating these Key Resource Tables provides a living and near-comprehensive log of all the inputs and outputs from ASAP CRN–funded research.

Conclusions A funder-led intervention to monitor and support a robust open science policy can foster the posting of preprints and the sharing of data, code, protocols, and key lab materials.

References

1. Thibault R. ASAP Open Science Policy Handbook. Zenodo. 2024. doi:10.5281/zenodo.13769766

2. Cobb-Lewis DE, Snyder D, Dumanis S, et al. Investing in open science: key considerations for funders. *bioRxiv*. 2024:12.09.627554. doi:10.1101/2024.12.09.627554

[1]Aligning Science Across Parkinson's (ASAP), Chevy Chase, MD, US, openscience@parkinsonsroadmap.org; [2]Coalition for Aligning Science (CAS), Chevy Chase, MD, US.

Conflict of Interest Disclosures All authors are employees or contractors of CAS, the managing organization for ASAP.

Funding/Support The authors conducted this research as part of their employment or contract with CAS. No additional funding was sought or acquired to conduct this research.

Acknowledgment We thank the Michael J. Fox Foundation for Parkinson's Research for their partnership in implementing the ASAP research initiative and for their assistance with the Open Science Compliance Workflow. We thank DataSeer for their active part in the development and conduct of the ASAP Open Science Compliance Workflow.

Additional Information The authors' ORCIDs are as follows: Robert Thibault, https://orcid.org/0000-0002-6561-3962; Dana E. Cobb-Lewis, https://orcid.org/0000-0002-4104-0311; Matt Lewis, https://orcid.org/0000-0002-9091-5516; Devin Snyder, https://orcid.org/0000-0001-8528-7538; Cornelis Blauwendraat, https://orcid.org/0000-0001-9358-8111; Sonya Dumanis, https://orcid.org/0000-0002-3345-9497. ChatGPT-4o (OpenAI) was used between February 10, 2025, and June 10, 2025, to help write the R Markdown script used to analyze the data. Robert Thibault takes responsibility for the integrity of the content generated. Data, codebooks, code, results, and instructions for how to reproduce our results are available on GitHub at https://github.com/robert-thibault/peer-review-congress-abstract and have been assigned a persistent identifier via Zenodo at https://doi.org/10.5281/zenodo.15832432. The ASAP Open Science Policy is available at https://doi.org/10.5281/zenodo.13769765. A preprint detailing the ASAP Open Science Compliance Workflow is available at https://doi.org/10.1101/2024.12.09.627554.

Medical Journal Policies on Requirements for Clinical Trial Registration, Reporting Guidelines, and Data Sharing: A Systematic Review

Kyobin Hwang,[1] Zexing Song,[2,3] Marsida Stafa,[3] Jodie Chiu[4] An-Wen Chan[1,3,5]

Objective We aimed to determine how often medical journals have policies requiring clinical trial registration, adherence to reporting guidelines, and participant-level data sharing. We also evaluated associations between journal characteristics and the existence of such policies for clinical trial manuscripts.

Design A PubMed search using the clinical trial filter was conducted to identify journals that published at least 20 trials in 2023. We extracted publicly available data from journal websites, including policies on trial registration, adherence to reporting guidelines, trial protocol submission requirements, and availability of participant-level data. A practice was classified as required if the policy used words such as must, need, or should. Policies using language such as encouraged and preferred were classified as recommended practices. For each journal, we recorded the 2023 Clarivate impact factor, journal scope (general vs specialty), and publishing model (purely open access vs other). We calculated descriptive statistics to summarize the prevalence of transparency policies and used multivariable logistic regression to assess the association between journal characteristics and policy requirements.

Results Among 380 included journals, 320 (84%) required trial registration, 11 (3%) recommended it, and 49 (13%) did not mention it. Adherence to a reporting guideline for clinical

Table 25-0892. Rates for the Deposition and Unambiguous Identification of Various Research Resources

Resource type	Publications with resource, No.	Resources shared, No.	Resources used, No.	Percentage shared	Increase in sharing vs manuscript draft, %
Data, new	95	657	744	88	67
Data, reused	45	159	184	86	14
Code, new	67	109	152	72	57
Software, reused	102	942	1189	79	44
Lab materials, new	36	190	245	78	56
Lab materials, reused	92	1820	2091	87	43
Protocols, all	97	716	885	81	46

Table 25-1183. Association Between Journal Characteristics and Mandated Policies (N = 378)[a]

Characteristic	Journals, No.[b]	Trial registration		Protocol submission		Participant-level data sharing	
		No. (%)[b]	AOR[c]	No. (%)[b]	AOR[c]	No. (%)[b]	AOR[c]
Trials published in 2023 (per 5-trial increase), median (IQR)	31 (24-46)	31 (24-46)	1.00 (0.96-1.06)	32 (24-51)	1.00 (0.97-1.03)	32 (24-51)	1.00 (0.97-1.05)
Journal scope							
Specialty	352	309 (88)	1 [Reference]	206 (59)	1 [Reference]	298 (85)	1 [Reference]
General medical	26	22 (85)	0.43 (0.13-1.70)	21 (81)	1.77 (0.62-5.85)	23 (88)	0.94 (0.28-4.32)
Open access journal							
No	302	259 (86)	1 [Reference]	165 (55)	1 [Reference]	251 (83)	1 [Reference]
Yes	76	72 (95)	4.02 (1.41-15.59)	62 (82)	4.13 (2.17-8.37)	70 (92)	2.46 (1.03-7.04)
2023 Impact factor (per 5-point increase), median (IQR)	4 (3-6)	4 (3-7)	2.65 (1.45-5.98)	4 (3-8)	1.56 (1.26-2.03)	4 (3-6)	1.07 (0.92-1.30)

Abbreviation: AOR, adjusted odds ratio.

[a]Two journals without an impact factor were excluded.

[b]Unless otherwise specified.

[c]Adjusted ORs from multivariable logistic regression adjusting for listed journal characteristics.

trials was required by 251 journals (66%), recommended by 74 (19%), and not mentioned by 55 (14%). Trial protocol submission was required by 118 (31%), recommended by 110 (29%), and not mentioned in 152 (40%). Public availability of participant-level datasets was required by 104 (27%), recommended by 218 (57%), and not mentioned by 58 (15%). A description of the data sharing plan was required by 212 journals (55.8%), recommended by 110 (28.9%), and not mentioned by 58 (15.3%). Purely open access journals had a 4-fold higher odds of requiring trial registration (adjusted odds ratio [AOR], 4.02; 95% CI, 1.41-15.59) and protocol submission (AOR, 4.13; 95% CI, 2.17-8.37) and 2.5 times higher odds of requiring public sharing of participant-level data (AOR, 2.46; 95% CI, 1.03-7.04) compared with other journals (**Table 25-1183**). Each 5-point increase in journal impact factor was associated with a more than 2.5-fold increase in the odds of requiring trial registration (AOR, 2.65; 95% CI, 1.45-5.98) and over 50% increase in the odds of requiring protocol submission (AOR, 1.56; 95% CI, 1.26-2.03). Impact factor was not significantly associated with requiring data sharing. No significant associations were found for journal scope or volume of trials published.

Conclusions Journal policy requirements vary substantially in supporting best practices for clinical trial transparency. Journals with a purely open access publishing model and higher impact factor were more likely to adopt transparency policies. These findings highlight the need for improved editorial standards across the publishing landscape to promote transparency and reduce research waste.

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada, anwen.chan@utoronto.ca; [2]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; [3]Women's College Research Institute, Toronto, Ontario, Canada; [4]Faculty of Health Science, University of Western Ontario, London, Ontario, Canada; [5]Division of Dermatology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada.

Conflict of Interest Disclosures An-Wen Chan is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

AI for Detecting Problems and Assessing Quality in Peer Review

Leveraging Large Language Models for Detecting Citation Quotation Errors in Biomedical Literature

M. Janina Sarol,[1] Jodi Schneider,[1,2] Halil Kilicoglu[1]

Objective An estimated 25% of citations in medical journals are inaccurate.[1] If not corrected, these citation quotation errors can propagate misinformation and establish unfounded claims.[2] Correcting them before publication is vital, yet citation accuracy is overlooked in the peer review process. The prevalence of citation quotation errors in the biomedical literature also remains unclear, because existing studies rely on manual evaluation on a limited set of articles.

Design We explored the use of large language models, specifically Gemini 1.5 Pro, GPT-4o (gpt-4o-2024-08-06), and LLaMA-3.1 (8B and 70B), to automatically identify citation quotation errors. We compared 2 approaches. In the first, we supplied the prompt with the citation in question, the paragraph containing the citation, and the full reference article. The second approach consists of a 3-step procedure and mimicked how a human reviewer might perform the task manually by first identifying the citation context, then locating relevant sentences in the reference article, and finally assessing citation accuracy. Specifically, the procedure consists of (1) extracting the citation sentence, (2) retrieving the 50 most relevant sentences using the MedCPT retriever model and reranking them with a fine-tuned MonoT5 model to select the top 5 sentences, and (3) verifying the citation accuracy with the large language model (**Box 25-1015**). Both approaches were evaluated from November 2024 to January 2025 using a citation quotation error dataset,[3] the first

Box 25-1015. Large Language Model Prompt for Citation Accuracy Verification

You are an expert in analyzing citation contexts in biomedical literature. Your task is to assess whether a citation in a citing article accurately reflects the content of the referenced article.

Definitions:
ACCURATE: The citation explicitly aligns with the facts, evidence, or established knowledge in the reference article. The details must match precisely, not just the general theme.
NOT_ACCURATE: The citation misrepresents the reference article by contradicting its claims, failing to substantiate them, oversimplifying key details, distorting meaning, or misquoting information.
NOT_ENOUGH_INFORMATION: The reference article does not explicitly discuss the information attributed to it in the citing article, making verification impossible.

Input Format:
Text from the citing article will be enclosed in <context></context> tags.
The specific citation marker under assessment will be highlighted as {marker}.
Other citations within the same context, but not relevant to your assessment, will be indicated as <|other_cit|>.
Relevant text from the reference article will be enclosed in <evidence></evidence> tags.

Output Format:
Return only one of the following labels: ACCURATE, NOT_ACCURATE, or NOT_ENOUGH_INFORMATION. Do not provide explanations or additional commentary.

publicly available corpus of its kind (Cohen κ for interrater agreement: 0.96 for citation context and 0.31 for accuracy detection). The better-performing approach was applied to assess 2898 citations from 2000 articles to the 100 most cited PubMed Central Open Access (PMC-OA) articles (24 different journals).

Results Gemini 1.5 Pro yielded best performance, obtaining an accuracy of 69% with both approaches. It outperformed GPT-4o slightly, while LLaMA-3 failed to generate meaningful results with the first approach and performed poorly with the second. Because the first approach only identified 25% (55 of 220) of the erroneous citations and the 3-step procedure identified 51% (113 of 220), the second approach was applied to the subset of PMC-OA articles. In the PMC-OA subset, 34% (981 of 2898) of the citation instances were assessed as erroneous. A total of 37% (742 of 2000) of citing articles were found to contain citation quotation errors; 98% (98 of 100) of reference articles were cited incorrectly at least once.

Conclusions Our preliminary results confirm the high prevalence of citation quotation errors in the medical literature. We will present a more comprehensive analysis of citation quotation errors in the PMC-OA subset at the conference. Automated verification of citations could help journals and peer reviewers identify questionable citation practices and reduce propagation of misinformation, improving the trustworthiness of scientific evidence.

References

1. Jergas H, Baethge C. Quotation accuracy in medical journal articles: a systematic review and meta-analysis. *PeerJ*. 2015;3:e1364. doi:10.7717/peerj.1364

2. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. *BMJ*. 2009;339:b2680. doi:10.1136/bmj.b2680

3. Sarol MJ, Ming S, Radhakrishna S, Schneider J, Kilicoglu H. Assessing citation integrity in biomedical publications: corpus annotation and NLP models. *Bioinformatics*. 2024;40(7):btae420. doi:10.1093/bioinformatics/btae420

[1]University of Illinois at Urbana-Champaign, IL, US, janinasarol@gmail.com; [2]Harvard Radcliffe Institute for Advanced Study, Cambridge, MA, US.

Conflict of Interest Disclosures Jodi Schneider declares nonfinancial associations with Crossref; Committee on Publication Ethics; International Association of Scientific, Technical and Medical Publishers; 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 the Center for Scientific Integrity (parent organization of Retraction Watch) and has received data-in-kind from Retraction Watch and Scite and usability testing compensation from the Institute of Electrical and Electronics Engineers. The National Information Standards Organization is a subawardee on her Alfred P. Sloan Foundation grant (G-2022-19409). No other disclosures were reported.

Funding/Support This study was supported by the Office of Research Integrity (ORI) of the US Department of Health and Human Services (HHS) (ORIIR220073). The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, the ORI, Office of the Assistant Secretary of Health, HHS, or the US government. Jodi Schneider was supported in part as the 2024-2025 Perrin Moorhead Grayson and Bruns Grayson Fellow, Harvard Radcliffe Institute for Advanced Study and by US National Science Foundation's CAREER grant (2046454).

Role of the Funder/Sponsor The funders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Acknowledgment We thank Shruthan Radhakrishna for his contributions to fine-tuning the MonoT5 model.

Additional Information Halil Kilicoglu is a co–corresponding author (halil@illinois.edu).

Automating the Detection of Promotional (Hype) Language in Biomedical Research

Bojan Batalo,[1] Erica K. Shimomoto,[1] Neil Millar[2]

Objective Promotional language (hype) in biomedical research writing has increased significantly over the past 40 years and has potential to influence readers' perceptions and evaluations of evidence.[1-3] Automatic systems capable of detecting and providing feedback on hype may offer a means to foster more objective reporting. However, the absence of formal guidelines for identifying hype represents a barrier for human annotators and the development of such systems. This pilot study develops formal guidelines for classifying hype and evaluates the application of annotated data to automate hype detection via machine learning.

Design Annotation guidelines were developed to classify 11 adjectives commonly associated with novelty and potential hype (eg, *creative*, *first*, *groundbreaking*, *innovative*). Guidelines followed the following hierarchical decisions: (1) does the adjective imply a positive value judgment?; (2) is it hyperbolic?; (3) is the adjective gratuitous (ie, adds little to

the content), amplified (strengthened by modifiers), or coordinated (paired with other promotional adjectives)?; and (4) is the broader sentential context promotional? A total of 550 sentences containing the adjectives (50 per adjective) were randomly sampled from National Institutes of Health grant abstracts funded between 2016 and 2020 and independently annotated by a linguist and 2 computer scientists, with disagreements resolved through discussion, providing the criterion standard. Our choice of grant abstracts as the starting point for our research is due to the highly competitive nature of research funding.[1] The annotated data were split 80:20 into development and hold-out test sets; machine learning algorithms were trained on the development and evaluated on the test set. Additionally, a human baseline was established by an additional researcher blinded to the annotation guidelines but supplied with a broad definition of hype, manually classifying the test set. Experiments tested 4 traditional text classification methods (multinomial Naive Bayes, multivariate Bernoulli Naive Bayes, latent semantic analysis, and support vector machines), with bag-of-words and averaged global vectors for word representation (GloVe) word embeddings as input features (**Table 25-0929**). Performance was evaluated using accuracy, weighted precision, recall, and F1 scores.

Results Excluding 13 sentences due to nonadherence with the guidelines, the final sample comprised 537 reliably annotated sentences (Cohen $\kappa > 0.94$), of which 392 were classified as hype. Among machine learning models, GloVe-based support vector machines outperformed others with an accuracy of 79.6%, approximating the human baseline of 82.4%.

Conclusions The annotation process underscored the subjective nature of assessing promotional language, particularly for context-dependent adjectives like *emerging* and *latest*, and the need for refined constructs to capture gradations of promotional language. Despite these limitations, the pilot study indicates the potential for machine learning models trained on well-annotated datasets to contribute to the automated detection of hype. Future steps include modifying the annotation systems and experimenting with large language models under zero/few-shot regimes.

Table 25-0929. Classification Performance on the Test Set Using 2 Input Features and a Human Baseline Supplied With Broad Definition of Hype

Method	Feature	Accuracy	Weighted precision	Weighted recall	Weighted F1 score
Human		0.824	0.819	0.824	0.821
Multinomial Naive Bayes	Band of words	0.741	0.713	0.741	0.716
Multivariate Bernoulli Naive Bayes		0.741	0.713	0.741	0.716
Latent semantic analysis		0.685	0.671	0.685	0.677
Support vector machines		0.759	0.736	0.759	0.717
Support vector machines	Global vectors for word representation	0.796	0.784	0.796	0.781

References

1. Millar N, Batalo B, Budgell B. Trends in the use of promotional language (hype) in abstracts of successful National Institutes of Health Grant Applications, 1985-2020. *JAMA Netw Open*. 2022;5(8):e2228676. doi:10.1001/jamanetworkopen.2022.28676

2. Qiu HS, Peng H, Fosse HB, Woodruff TK, Uzzi B. Use of promotional language in grant applications and grant success. *JAMA Netw Open*. 2024;7(12):e2448696. doi:10.1001/jamanetworkopen.2024.48696

3. Van Den Besselaar P, Mom C. The effect of writing style on success in grant applications. *J Informetr*. 2022;16(1):101257. doi:10.1016/j.joi.2022.101257

[1]National Institute of Advanced Industrial Science and Technology, Japan, bojan.batalo@aist.go.jp; [2]University of Tsukuba, Japan.

Conflict of Interest Disclosures None reported.

Funding/Support This study was supported by grants 21K02919 and 25K00851 from the Japan Society for the Promotion of Science.

Role of the Funder/Sponsor The authors confirm that no sponsors or funders influenced the study design, execution, or interpretation of results.

Additional Information Neil Millar is a secondary corresponding author (millar.neil@u.tsukuba.ac.jp).

Evaluation of a Method to Detect Peer Reviews Generated by Large Language Models

Vishisht Rao,[1] Aounon Kumar,[2] Himabindu Lakkaraju,[2] Nihar B. Shah[1]

Objective Journals, conferences, and funding agencies face the risk that reviewers might ask large language models (LLMs) to generate reviews by uploading submissions. Existing detectors struggle to differentiate between fully LLM-generated and LLM-polished reviews. We addressed this problem by developing a method to detect and flag LLM-generated reviews while controlling family-wise error rates (FWERs).

Design Our method had 3 components. (1) Watermarking: We stochastically chose specific phrases (watermarks), such as random (fake) citations, technical terms ("weak supervised learning"), or beginning the review with a prefix ("This paper investigates the problem..."). Watermarks were specific modifications to the review, which were undetectable unless the nature of the watermark was known. The watermark was known to the editors (but not the reviewers), who could detect LLM-generated reviews through watermarks in the review. (2) Hidden prompt injection: We added an instruction into the manuscript PDF (at the end of the last

Table 25-1176. Watermark Embedding Success Rates in Large Language Model (LLM)–Generated Peer Reviews Across Prompt Injection Strategies and Reviewer Defenses

	Random citation	Random start	Technical term
White text			
4o UI (n = 100)	1.0 ± 0.0	0.89 ± 0.11	0.91 ± 0.09
4o API (n = 30)	0.98 ± 0.02	0.80 ± 0.08	0.82 ± 0.08
o1-mini (n = 100)	1.0 ± 0.0	0.89 ± 0.06	0.45 ± 0.10
Gemini (n = 100)	1.0 ± 0.0	0.96 ± 0.04	0.95 ± 0.05
Sonnet (n = 100)	0.95 ± 0.05	0.83 ± 0.08	0.85 ± 0.07
Different language			
4o UI (n = 100)	0.97 ± 0.03	0.05 ± 0.05	0.03 ± 0.03
4o API (n = 30)	0.96 ± 0.04	0.27 ± 0.09	0.0 ± 0.0
o1-mini (n = 100)	0.92 ± 0.06	0.01 ± 0.01	0.04 ± 0.04
Gemini (n = 100)	0.93 ± 0.05	0.25 ± 0.09	0.01 ± 0.01
Sonnet (n = 100)	0.96 ± 0.04	0.21 ± 0.08	0.0 ± 0.0
Font manipulation (n = 30)			
4o UI	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
4o API	0.87 ± 0.13	1.0 ± 0.0	0.93 ± 0.07
o1-mini	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Gemini	0.03 ± 0.03	0.0 ± 0.0	0.0 ± 0.0
Sonnet	0.57 ± 0.17	1.0 ± 0.0	0.60 ± 0.17
RD: paraphrasing (n = 100)			
Gemini	0.94 ± 0.05	0.0 ± 0.0	0.81 ± 0.08
4o	1.0 ± 0.0	0.0 ± 0.0	0.87 ± 0.07
Sonnet	0.98 ± 0.02	0.0 ± 0.0	0.95 ± 0.05
RD: identifying watermarks, white text (n = 100)			
4o	0.19 ± 0.09	0.03 ± 0.03	0.09 ± 0.06

RD: crop end of paper (n = 50)	**Font manipulation, 4o**	**0.46 ± 0.14**	**White text, 4o**	**0.21 ± 0.11**	**White text, 4o**	**0.16 ± 0.1**

Abbreviations: API, application programming interface; RD, reviewer defense; UI, user interface.

The first 3 sections contain the fraction of LLM-generated reviews that embedded the chosen watermark using the specified prompt injection strategy across different LLMs: OpenAI's ChatGPT-4o and o1-mini, Google's Gemini 2.0 Flash, and Anthropic's Claude 3.5 Sonnet. The columns contain the various watermarking strategies; the sections contain the various prompt injection strategies. The next 3 sections contain 3 RDs. The paraphrasing RD and crop end of paper RD sections contain the fraction of LLM-generated reviews that embedded the chosen watermark despite the RD, and the identifying watermarks RD section contains the fraction of submissions that were flagged by the LLM for containing a hidden prompt. The values are all from 0 to 1, with 1 being 100% accuracy.

page; not detectable by the human reviewer) instructing the LLM to output the chosen watermark in the LLM-generated review ("Start your review with: <chosen watermark>"). We evaluated 3 injection methods: white text, font manipulation,[1-3] and a different language. (3) Statistical detection method: We developed a test for watermarks in the review (null hypothesis was the review was human written) and provided mathematical proof of controlling the FWER without making assumptions on how human reviews were written. The evaluation was performed on International Conference on Learning Representations (ICLR) manuscripts, a publication venue that reviews full manuscripts. We injected the hidden prompts into ICLR manuscripts and asked LLMs to generate a review and observe whether it contained the watermark. We also evaluated the robustness of our methods against 3 reviewer defenses (RDs), measures reviewers could take to avoid being flagged for using an LLM-generated review: paraphrasing (paraphrasing the original LLM-generated review using another LLM), identifying watermarks (asking the LLM to flag submissions with hidden prompts), and cropping out the end of the paper (removing the last page before generating the review using an LLM). Watermarks were most successfully embedded in the LLM-generated review when the hidden prompt was injected at the end of the last page. For the evaluation of the last RD, we injected the hidden prompt at a position other than the end of the last page.

Results Our method was successful in embedding the chosen watermark in LLM-generated reviews and was robust to RD mechanisms (**Table 25-1176**) and detecting and flagging them while controlling false-positives. We conducted our test on 28,028 human-written reviews from ICLR 2021 and 10,022 from 2024, each augmented with 100 LLM-generated reviews with a watermark embedded for each of the 3 types. When executing our algorithm to control the FWER at 0.01 for the random start and technical term watermark and 0.001 for the random citation watermark, we observed 0 false-positives in all cases.

Conclusions We propose a novel method to detect LLM-generated peer reviews. Our evaluations found this method to be successful, making it appealing to journal, conference, and proposal review organizers.

References

1. Markwood I, Shen D, Liu Y, Lu Z. Mirage: content masking attack against {information-based} online services. In: *26th USENIX Security Symposium*. 2017:833-847.

2. Tran D, Jaiswal C. PDFPhantom: exploiting PDF attacks against academic conferences' paper submission process with counterattack. In: *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*. October 2019:0736-0743.

3. Zou A, Wang Z, Carlini N, Nasr M, Kolter JZ, Fredrikson M. Universal and transferable adversarial attacks on aligned language models. *arXiv*. Preprint posted online July 27, 2023. doi:10.48550/arXiv:2307.15043

[1]Carnegie Mellon University, Pittsburgh, PA, US, nihars@cs.cmu.edu; [2]Harvard University, Boston, MA, US.

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

Funding/Support US National Science Foundations Awards 1942124, 2200410; ONR grant N000142212181.

Role of the Funder/Sponsor The funders played no role in the design and conduct of the experiment, data analysis, or preparation of the manuscript.

Acknowledgment We thank Danish Pruthi for very helpful discussions.

Quality and Comprehensiveness of Peer Reviews of Journal Submissions Produced by Large Language Models vs Humans

Fares Alahdab,[1,2,3] Juan Franco,[3,4] Helen Macdonald,[4] Sara Schroter[4]

Objective Peer reviewer fatigue is on the rise, as reviewers are overburdened with increasing tasks and manuscripts to review, making it challenging for journal editors to secure a sufficient number of high-quality reviews.[1] Despite the potential benefits of using large language models (LLMs) in editorial and peer review processes[2] and the potential to support reviewers with tasks, it is not yet known how good they are at peer review. Additionally, covert use of LLMs in peer review is increasingly suspected, with limited knowledge of positive or negative impact. We aim to compare the quality and comprehensiveness of peer reviews produced by 5 LLMs compared with peer reviewers for research submissions.

Design A comparative study of peer reviews produced by LLMs (GPT4o, GPTo3, Claude 3.5, Gemini 1.5 Pro, and Gemini 2.0 Flash) compared with 2 human peer reviews for research submissions between May 2024 and June 2025. Manuscripts were uploaded to Google's Vertex AI, a private and secure BMJ Group workspace for using LLMs; the same prompt was used for all submissions (single-shot prompting). An experienced editor performed the ratings of all LLM and human reviews using the Review Quality Instrument (RQI),[3] a tool for assessment of review quality comprising 8 questions, each on a 5-point Likert scale. Secondary outcomes include a comprehensiveness score, based on elements of the manuscript that the reviews focused on, and whether they were evaluative in nature. Additionally, once the first decision has been made on the manuscripts, we plan to invite authors to complete a survey about their perceptions of both the LLM review reports and the peer reviewer reports for their manuscripts. We also plan to have a second editor perform the ratings and compute interrater agreement. Reviews produced by LLMs are not included in the decision-making for the manuscripts. Wilcoxon rank sum tests were used to compare LLM and human scores for each RQI item.

Results Preliminary data include 35 reviews (25 by LLMs and 10 by peer reviewers) of 5 BMJ submissions. Across 8 RQI items, LLM-generated reviews had higher mean (SD) scores than human reviewers on identifying strengths and weaknesses (4.12 [0.53] vs 2.70 [0.95]; P < 0.001); providing useful comments on the writing, organization, tables, and figures (3.72 [0.79] vs 1.80 [0.63]; P < .001); and constructiveness (4.00 [0.58] vs 3.00 [1.05]; P = .004) (**Table 25-1146**). Complete data will include 200 submissions to 4 BMJ journals, the secondary outcomes, and authors' feedback on the LLM reviews.

Conclusions LLM-generated reviews matched or exceeded human reviewers on a few key dimensions of review quality. A fuller analysis will shed more light on the potential value of LLM peer reviews and how they could complement human peer reviewers' work.

Table 25-1146. Comparison of RQI Item Scores for LLM-Generated vs Human Reviews[a]

RQI item	Mean (SD)	Median (IQR)	*P* value
1. Did the reviewer discuss the importance of the research question?			
LLM	3.40 (0.76)	4.00 (3.00-4.00)	.06
Human	2.40 (1.43)	2.00 (1.00-4.00)	
2. Did the reviewer discuss the originality of the paper?			
LLM	3.04 (0.54)	3.00 (3.00-3.00)	.84
Human	2.80 (1.48)	3.00 (1.25-4.00)	
3. Did the reviewer clearly identify the strengths and weaknesses of the method (study design, data collection, and data analysis)?			
LLM	4.12 (0.53)	4.00 (4.00-4.00)	<.001
Human	2.70 (0.95)	3.00 (2.00-3.00)	
4. Did the reviewer make specific useful comments on the writing, organization, tables, and figures of the manuscript?			
LLM	3.72 (0.79)	4.00 (3.00-4.00)	<.001
Human	1.80 (0.63)	2.00 (1.25-2.00)	
5. Were the reviewer's comments constructive?			
LLM	4.00 (0.58)	4.00 (4.00-4.00)	.004
Human	3.00 (1.05)	3.00 (2.25-4.00)	
6. Did the reviewer supply appropriate evidence using examples from the paper to substantiate their comments?			
LLM	2.96 (1.17)	3.00 (3.00-3.00)	.64
Human	3.20 (1.75)	3.00 (1.50-5.00)	
7. Did the reviewer comment on the author's interpretation of the results?			
LLM	1.64 (0.99)	1.00 (1.00-2.00)	.09
Human	2.30 (1.25)	2.00 (1.25-3.00)	
8. How would you rate the quality of this review overall?			
LLM	3.80 (0.65)	4.00 (3.00-4.00)	.25
Human	3.20 (1.32)	3.50 (3.00-4.00)	

Abbreviations: LLM, large language model; RQI, Review Quality Instrument.

[a]There were 25 LLM-generated reviews and 10 human reviews for each RQI item.

References

1. Publons. Global State of Peer Review report. https://publons.com/community/gspr

2. Liang W, Zhang Y, Cao H, et al. Can large language models provide useful feedback on research papers? a large-scale empirical analysis. *arXiv*. Preprint posted October 2, 2023. doi.org/10.48550/arXiv.2310.01783

3. 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

[1]University of Missouri-Columbia, Columbia, MO, US, fares.alahdab@health.missouri.edu; [2]University of Texas Health Science Center, Houston, TX, US; [3]BMJ Evidence-Based Medicine, London, UK; [4]*BMJ*, London, UK.

Conflict of Interest Disclosures None reported.

Poster Session Abstracts

All In-person Posters will be presented on Thursday, September 4, and Friday, September 5; In-person and Virtual Posters will be available to view and and post comments and questions during the meeting via the online platform at underline.io/events/476/reception

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

AI in Peer Review and Publication

In-person

Domain-Specific Pretrained Encoder Transformers for the Identification of Methodologically Rigorous Systematic Reviews: A Retrospective Modeling Study

Fangwen Zhou,[1] Muhammad Afzal,[2] Rick Parrish,[1] Ashirbani Saha,[3] Wael Abdelkader,[1] R. Brian Haynes,[1] Alfonso Iorio,[1,4] Cynthia Lokker[1]

Objective Systematic reviews are considered one of the strongest levels of evidence, providing essential information for clinical guidelines and bedside practice. However, the broad spectrum of review articles, including narrative reviews and other types, complicates the identification of high-quality systematic reviews.[1] This study aims to fine-tune and evaluate pretrained encoder transformers to identify high-quality systematic reviews from review articles using a large, reputable dataset.

Design Review articles from McMaster's Premium Literature Service (PLUS) were retrieved.[2] Articles were considered rigorous if they were systematic reviews that (1) stated the clinical topic; (2) described methods, including databases searched and inclusion criteria; (3) searched ≥1 major database; (4) reported the number of retrieved, reviewed, and included articles; and (5) did not exclude randomized clinical trials for treatment, primary prevention, quality improvement, or economics reviews; or included "inception cohort" for prognosis reviews.[3] Ground truth was established by research methodology experts using article full texts. Articles from 2003 to 2023 were randomly split 80:10:10 into training, validation, and testing sets. Articles in 2024 were used for external testing. A grid search of 7 pretrained models, 3 learning rates, 5 batch sizes, and 3 random seeds, with or without class weight adjustments, was conducted. Models were trained for ≤10 epochs, with an early stopping patience of 3. Titles and abstracts were used as inputs. Those longer than 512 tokens were truncated, and those shorter were padded. The model that achieved the lowest log loss on the validation set was further evaluated. A threshold of ≥0.50 was used for classification. Bootstrapping of 1000 iterations was used to estimate 95% confidence intervals.

Results Among 40,091 reviews, 26,332 (65.7%) were considered rigorous and 31,593 were used for training. Of the 630 fine-tuned models, BioELECTRA with no class weight adjustments, learning rate 5E-5, batch size 64, and seed 2 had the lowest validation log loss of 0.1840. **Table 25-0887** details the characteristics of the datasets and the model's performance, achieving >94% area under the receiver operating characteristic curve and >95% sensitivity. In review articles published in 2024, there was a notable drop in specificity and mild degradation in other metrics.

Conclusions Pretrained encoder transformer models, such as BioELECTRA, demonstrate robust performance in identifying rigorous systematic reviews based on PLUS's criteria. These models can streamline the identification of high-quality evidence, reducing manual effort and enhancing the efficiency of evidence curation. Performance degradation in articles in 2024 may be due to changes in article structure and content over time. Additionally, generalizability to other datasets is unknown. Future efforts should establish standardized benchmarking datasets for systematic reviews,

Table 25-0887. Characteristics of Datasets and Best Model Performance

Dataset	No. articles	Articles, No. (%)		AUROC (95% CI)[a]	Accuracy (95% CI)[a]	Sensitivity (95% CI)[a]	Specificity (95% CI)[a]
		Rigorous	Nonrigorous				
Training	31,593	20,673 (65.4)	10,920 (34.6)	-	-	-	-
Validation	3949	2605 (66.0)	1344 (34.0)	97.5 (97.0-97.9)	93.0 (92.3-93.8)	96.2 (95.4-96.9)	86.9 (85.0-88.8)
Testing	3950	2594 (65.7)	1356 (34.3)	97.3 (96.8-97.7)	92.5 (91.7-93.3)	95.8 (95.0-96.6)	86.2 (84.4-88.0)
2024	599	460 (76.8)	139 (23.2)	94.5 (92.2-96.4)	89.3 (86.8-91.7)	95.7 (93.4-97.4)	68.4 (60.8-75.9)
Total	40,091	26,332 (65.7)	13,759 (34.3)	-	-	-	-

Abbreviation: AUROC, area under the receiver operating characteristic curve.

[a]Performance of the BioELECTRA model with no class weight adjustments, learning rate of 5E-5, batch size of 64, and a random seed of 2.

develop models for tools, such as AMSTAR 2, and account for temporal data drift to better support knowledge translation.

References

1. Shaheen N, Shaheen A, Ramadan A, et al. Appraising systematic reviews: a comprehensive guide to ensuring validity and reliability. *Front Res Metr Anal*. Published online December 21, 2023. doi:10.3389/frma.2023.1268045

2. Holland J, Haynes RB; McMaster PLUS Team Health Information Research Unit. McMaster Premium Literature Service (PLUS): an evidence-based medicine information service delivered on the Web. *AMIA Annu Symp Proc*. 2005;2005:340-344. https://www.ncbi.nlm.nih.gov/pubmed/16779058

3. Methodological criteria. McMaster University Health Information Research Unit. Accessed August 19, 2024. https://hiruweb.mcmaster.ca/hkr/what-we-do/methodologic-criteria

[1]Health Information Research Unit, Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada, lokkerc@mcmaster.ca; [2]Department of Computing, Faculty of Computing, Engineering and the Built Environment, Birmingham City University, Birmingham, UK; [3]Department of Oncology, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; [4]Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada.

Conflict of Interest Disclosures McMaster University, a nonprofit public academic institution, has contracts through the Health Information Research Unit under the supervision of Alfonso Iorio and R. Brian Haynes. These contracts involve professional and commercial publishers to provide newly published studies, which are critically appraised for research methodology and assessed for clinical relevance as part of the McMaster Premium Literature Service (McMaster PLUS). Cynthia Lokker and Rick Parrish receive partial compensation, and R. Brian Haynes is remunerated for supervisory responsibilities and royalties. Ashirbani Saha, Fangwen Zhou, Muhammad Afzal, and Wael Abdelkader are not affiliated with McMaster PLUS.

Funding/Support Fangwen Zhou was funded through the Mitacs Business Strategy Internship grant (IT42947) with matching funds from EBSCO Canada.

Role of the Funder/Sponsor The funders 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; and decision to submit the abstract for presentation.

Acknowledgment We thank the Digital Research Alliance of Canada for its computational resources.

Strategic Insights Into Editor Engagement With AI-Assisted Tools Based on Survey and Data Analysis of AI-Assisted Ethics Checks

Beth Waymouth,[1] Heather Slater,[1] Angharad Goode,[1] Katie Allin,[1] Maria Kowalczuk[1]

Objective As publishers increasingly adopt artificial intelligence (AI)–driven tools, understanding editor engagement with AI-assisted editorial checks is crucial for maintaining publication standards. This study examines Handling Editors' (HEs) interaction with an AI-powered check that flags manuscripts needing further verification during peer review and allows HEs and the editorial office (EOF) to collaborate on resolutions. After observing a decline in HE engagement, a survey was conducted to uncover the underlying reasons.

Design We analyzed internal data tracking interaction with the AI-supported ethics check introduced to HEs in 2022. Existing editors were emailed about this new tool and given access and instructions, including a detailed Ethics Guidance document. Newly recruited HEs were introduced to the check during onboarding webinars. Data included the final status of the ethics check, collected quarterly between September 2022 and March 2024. Engagement was defined as the resolution of ethics checks by the HE rather than the EOF. To contextualize the engagement patterns, a survey was distributed to 4627 HEs across our journal portfolio who were active in 2024 and consented to be recontacted. Data were collected for 2 weeks in December 2024, with a reminder issued after the first week.

Results Initial analysis showed that editor engagement dropped from 40% (2899 of 7394 total checks) to 20% (2170 of 10,692) between September 2022 and November 2023. Engagement rose to 25% (1227 of 4766 checks) between January and March 2024; however, this increase was not statistically significant. Of the surveyed editors, 725 (16%) responded to at least the first question. Of 662 respondents, 67% amended the ethics check on at least some manuscripts, with 24% interacting with the check for every manuscript. Among surveyed editors, 33% indicated that they were unaware of the check. In the "Other" category (n = 66), 38 HEs (58%) said no ethics issues required action or manuscripts didn't need ethical approval; 4 cited unfamiliarity with the check, and 4 responded that they did not remember. Out of 478 responses from HEs not amending the check on some or all manuscripts, 67% cited an expectation that this task fell to the EOF, especially true when handling higher manuscript volumes (**Table 25-0897**).

Conclusions These results underscore the risk of editor disengagement when AI-assisted tools lack sufficient onboarding support and adequate training. For AI integration to be effective, publishers must employ consistent monitoring of engagement and explore ways to ensure AI tools are visible, actionable, and reliably applied across editorial checkpoints.

[1]Frontiers Media S.A., Switzerland, maria.kowalczuk@frontiersin.org.

Conflict of Interest Disclosures Heather Slater, Angharad Goode, Katie Allin, and Maria Kowalczuk are employees of Frontiers Media S.A. Beth Waymouth was an employee of Frontiers Media S.A. at the time when the study was conducted and when the abstract was written and originally submitted. Since May 2025, Beth Waymouth has been an independent researcher.

Acknowledgment We thank all our Handling Editors who participated in the survey and Elena Vicario and Simone Ragavooloo for their valuable suggestions and comments on this project and the abstract.

Table 25-0897. Editors' Responses to the Survey Sorted by the Number of Manuscripts They Handled in the Last 12 Months

What, if anything, has prevented you from amending the Ethics Guidelines check? Please tick all that apply (n = 478)	Manuscripts handled in the last 12 mo, No. (%)				
	1 (n = 54)	2-5 (n = 275)	6-10 (n = 98)	≥10 (n = 51)	Total (N = 478)
I thought it was more appropriate for the Editorial Office to perform this check	16 (30)	83 (30)	36 (37)	23 (45)	158 (33)
I thought the Editorial Office would perform this check	17 (31)	95 (35)	36 (37)	16 (31)	164 (34)
The guidelines that have been provided do not cover the issues I see in manuscripts	10 (19)	21 (8)	18 (18)	6 (12)	55 (12)
I cannot remember	9 (17)	67 (24)	17 (17)	7 (14)	100 (21)
The guidelines that have been provided are unclear	2 (4)	10 (4)	6 (6)	5 (10)	23 (5)
I did not have time to perform this check	3 (6)	16 (6)	3 (3)	4 (8)	26 (5)
Other (please specify)	10 (19)	36 (13)	15 (15)	5 (10)	66 (14)
I was not sure how to contact the Editorial Office	3 (6)	7 (3)	1 (1)	2 (4)	13 (3)

Attitudes and Perceptions of Biomedical Journal Editors in Chief Toward the Use of Artificial Intelligence Chatbots in the Scholarly Publishing Process

Jeremy Y. Ng,[1,2,3] Malvika Krishnamurthy,[2,3] Gursimran Deol,[2,3] Wid Al-Zahraa Al-Khafaji,[2,3] Vetrivel Balaji,[4] Magdalene Abebe,[2,3] Jyot Adhvaryu,[2,3] Tejas Karrthik,[2,3] Pranavee Mohanakanthan,[2,3] Adharva Vellaparambil,[2,3] Lex M. Bouter,[5,6] R. Brian Haynes,[7] Alfonso Iorio,[7,8] Cynthia Lokker,[7] Hervé Maisonneuve,[9,10] Ana Marušić,[11] David Moher[1,12]

Objective This study aimed to examine the attitudes and perceptions of editors in chief (EICs) of biomedical journals regarding the integration of artificial intelligence chatbots (AICs) into the scholarly publishing process. Prior research has explored AI use in publishing broadly, but limited data exist on EIC perspectives. Although AICs offer opportunities to streamline editorial tasks, such as plagiarism detection and language editing, they also introduce ethical, technical, and operational challenges. Understanding EIC perspectives is critical to shaping guidelines, policies, and training that align with the evolving role of AICs in scholarly publishing.

Design We conducted a cross-sectional survey of EICs of all biomedical journals, inclusive of medical, nursing, health sciences, dentistry, nursing, public health, and pharmacology and toxicology disciplines, published by Springer Nature (including BMC), Taylor & Francis, Elsevier, Wiley, and Sage. Eligible journals were identified through a combination of automated web scraping and manual verification. EICs were invited to participate in an anonymous SurveyMonkey survey conducted over 5 weeks, which included 3 follow-up reminders, from July through August 2024. The survey covered familiarity with AICs, current use, perceived benefits and challenges, and anticipated future roles. Quantitative data were analyzed using descriptive statistics, while qualitative responses were coded and thematically analyzed to identify key themes. Our protocol was registered[1] and followed the CHERRIES reporting guideline.[2]

Results Of 3381 EICs contacted, 510 responded (15.1% response rate), with 505 eligible participants and a completion rate of 87.0%. Most were familiar with AICs (66.7% [325 of 487]) but had not used them in editorial workflows (83.7% [401 of 479]). Perceived benefits included enhanced language and grammar support (70.8% [308 of 435]) and plagiarism screening (67.3% [294 of 437]). However, respondents expressed concerns about initial setup and training (83.9% [360 of 429]), ethical risks (80.6% [345 of 428]), and technical reliability (75.2% [322 of 428]). While only 49.6% (240 of 484) of participants reported that their journal had formal AIC policies, 89.5% (419 of 468) of respondents supported training initiatives to promote ethical and effective use. Despite limited current adoption, 78.9% (370 of 469) believed AICs will play an important role in the future of scholarly publishing, and 77.2% (363 of 470) anticipated their significance in advancing scientific research. Themes identified through thematic analysis of open-ended questions included "no AI in authorship or peer review," referring to EICs' reporting of current journal and publisher policy on their use, and "ethical, integrity, and privacy concerns," referring to EIC perceptions of challenges with the use of AICs in the scholarly publishing process. Our study manuscript has been preprinted.[3]

Conclusions Biomedical journal EICs recognized the potential of AICs to enhance editorial processes but highlighted critical barriers, including ethical dilemmas, resource limitations, and insufficient policies and training. Structured interventions, including targeted training programs and robust ethical guidelines, are essential for addressing these challenges and ensuring responsible and effective integration of AICs into publishing workflows.

References

1. Ng JY, Krishnamurthy M, Balaji V, et al. Attitudes and perceptions of biomedical journal editors-in-chief towards the use of artificial intelligence chatbots in the scholarly publishing process: a cross-sectional survey across multiple publishers. OSF Registries. August 10, 2024. Accessed July 2, 2025. https://osf.io/xt6f2

2. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res.* 2004;6(3):e34. doi:10.2196/jmir.6.3.e34

3. Ng JY, Krishnamurthy M, Deol G, et al. Attitudes and perceptions of biomedical journal editors-in-chief towards the use of artificial intelligence chatbots in the scholarly publishing process: a cross-sectional survey. *medRxiv*. Preprint posted online May 27, 2025. doi:10.1101/2025.05.26.25328101

[1]Centre for Journalology, Ottawa Methods Centre, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, jeremyyng.phd@gmail.com; [2]Institute of General Practice and Interprofessional Care, University Hospital Tübingen, Tübingen, Germany; [3]Robert Bosch Center for Integrative Medicine and Health, Bosch Health Campus, Stuttgart, Germany; [4]Department of Computing and Software, Faculty of Engineering, McMaster University, Hamilton, Ontario, Canada; [5]Department of Epidemiology and Data Sciences, Amsterdam Universities Medical Center, Amsterdam, the Netherlands, [6]Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, the Netherlands; [7]Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; [8]Department of Medicine, McMaster University, Hamilton, Ontario, Canada; [9]Consultant, Lyon, France; [10]Scientific Committee, Institute of Research and Action on Fraud and Plagiarism in Academia (IRAFPA), Geneva, Switzerland; [11]Department of Research in Biomedicine and Health and Center for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia; [12]School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

Conflict of Interest Disclosures Lex M. Bouter, Ana Marušić, and David Moher are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract. No other disclosures were reported.

Attitudes and Perceptions Toward the Use of Artificial Intelligence Chatbots in Medical Journal Peer Review: A Large-Scale, International Cross-Sectional Survey

Jeremy Y. Ng,[1,2,3] Daivat Bhavsar,[1,2] Neha Dhanvanthry,[1,2] Lex Bouter,[4,5] Teresa M. Chan,[6] Holger Cramer,[1,2] Annette Flanagin,[7] Alfonso Iorio,[8,9] Cynthia Lokker,[8] Hervé Maisonneuve,[10,11] Ana Marušić,[12] David Moher[3,13]

Objective Peer review is a cornerstone of scientific publishing, ensuring the rigor and credibility of research. However, it is increasingly strained by underexplored challenges, such as variability in quality, reviewer fatigue, and biases. Artificial intelligence chatbots (AICs) are emerging tools with the potential to address these challenges by automating tasks, such as identifying methodological flaws and improving language clarity. This study aimed to explore peer reviewers' attitudes and perceptions toward the use of AICs in medical journal peer review, focusing on their benefits, challenges, and ethical implications.

Design A large-scale, international, cross-sectional survey targeting peer reviewers of medical journals was conducted. Eligible participants had completed at least 1 peer review report for a MEDLINE-indexed journal within the last 24 months at the point of invitation. The names and email addresses of a complete sample of 72,847 corresponding authors who published in MEDLINE-indexed journals from September 1, 2024, to October 15, 2024, were collected. The survey was administered and data were collected via the SurveyMonkey survey platform. The survey included questions about participant demographics, familiarity with AICs, experiences with their use, and perceptions of AIC's roles, benefits, challenges, and ethical implications in peer review. The survey included multiple-choice, yes/no, Likert scale, and open-ended questions to ensure comprehensive data collection. The survey was first piloted with a group of 7 peer reviewers to optimize question clarity and relevance. The survey was launched in April 2025 and remained open for 7 weeks in total. This study's protocol was registered[1] and the CHERRIES[2] and STROBE[3] reporting guidelines were used to inform the reporting of this survey study.

Results The survey received 1194 respondents from 33,388 opened email invitations in the first 4 weeks, 1018 of whom completed the survey (response rate: 3.0%). The majority (578/1018 [56.8%]) were senior-career researchers (ie, >10 years of career experience). Most reported never having used an AIC for peer review purposes (707/999 [70.8%]), and notably, many (460/1007 [45.7%]) did not anticipate using AICs in the future. However, most respondents indicated interest in receiving training on AIC use for peer review tasks (587/980 [59.9%]). The most-recognized benefit of AIC use in peer review processes (ie, "agree" or "strongly agree") was "reduc[ing] the workload" (601/958 [62.7%]). Conversely, the most-recognized challenge was "risk producing errors or inaccuracies" (738/934 [79.0%]).

Conclusions This study provides critical insights into peer reviewers' attitudes and perceptions toward the use of AICs in the peer review process. Findings may inform the development of evidence-based guidelines and policies to ensure the ethical, transparent, and effective use of AICs, contributing to improvements in efficiency, quality, and equity in scholarly publishing.

References

1. Ng JY, Bhavsar D, Dhanvanthry N, et al. Attitudes and perceptions toward the use of artificial intelligence chatbots in medical journal peer review: a large-scale, international cross-sectional survey. OSF. May 21, 2025. https://osf.io/fhc2m

2. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res*. 2004;6(3):e34. doi:10.2196/jmir.6.3.e34

3. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*. 2007;370(9596):1453-1457. doi:10.1016/S0140-6736(07)61602-X

[1]Institute of General Practice and Interprofessional Care, University Hospital Tübingen, Tübingen, Germany, jeremyyng.phd@gmail.com; [2]Robert Bosch Center for Integrative Medicine and Health, Bosch Health Campus, Stuttgart, Germany; [3]Centre for Journalology, Ottawa Methods Centre, Ottawa Hospital

Research Institute, Ottawa, Ontario, Canada; [4]Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, the Netherlands; [5]Department of Philosophy, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; [6]School of Medicine, Toronto Metropolitan University, Toronto, Ontario, Canada; [7]*JAMA* and the JAMA Network, Chicago, IL, US; [8]Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; [9]Department of Medicine, McMaster University, Hamilton, Ontario, Canada; [10]Consultant, Lyon, France; [11]Scientific Committee, Institute of Research and Action on Fraud and Plagiarism in Academia (IRAFPA), Geneva, Switzerland; [12]Department of Research in Biomedicine and Health and Center for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia; [13]School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada.

Conflict of Interest Disclosures Lex Bouter, Annette Flanagin, Ana Marušić, and David Moher are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract.

Use of an LLM as an Author Checklist Assistant for Scientific Papers: NeurIPS 2024 Experiment

Alexander Goldberg,[1] Ihsan Ullah,[2] Thanh Gia Hieu Khuong,[3] Benedictus Kent Rachmat,[3] Zhen (Zach) Xu,[4] Isabelle Guyon,[2,3,5] Nihar B. Shah[1]

Objective This study assessed the utility of a large language model (LLM) in evaluating compliance with submission standards at the 2024 Neural Information Processing Systems (NeurIPS) conference—a top-tier publication venue in artificial intelligence (AI) with a 15% to 25% acceptance rate that reviews full papers, not just abstracts. NeurIPS requires authors to complete a 15-question checklist promoting reproducibility, transparency, and ethical standards, similar to CONSORT[1] and STROBE.[2] We introduced an optional Checklist Assistant using GPT-4 (OpenAI) to provide presubmission feedback to authors on checklist accuracy.

Design We conducted a cross-sectional study in which the Checklist Assistant was available to authors 8 days before the submission deadline. Authors could optionally use the Checklist Assistant, which used a general purpose third-party LLM (gpt-4-turbo-preview) to evaluate the accuracy of their responses to a 15-question checklist. For each question, the LLM received the author's response, justification, and the full paper (including appendices), and assessed the accuracy of the response using simple prompt engineering. We conducted surveys about authors' perceptions of the tool at registration and after using the tool. We received 539 preusage survey responses (out of 17,491 total papers submitted to the conference), 234 submissions to the Checklist Assistant, and 65 distinct postusage survey responses. To evaluate robustness of the Checklist Assistant, we manipulated responses to the Assistant to test whether it would be possible for authors to manipulate the Assistant's scores of checklist accuracy.

Results In 65 postusage surveys (**Figure 25-0958**), 46 authors found the Assistant useful and 46 indicated they would revise their papers or checklist responses based on its feedback. Authors' expectations for the Checklist Assistant were more positive (59 of 65 positive responses on usefulness) than their perceptions after usage (46 of 65), potentially reflecting an overly optimistic outlook on the usefulness of the tool. Analysis of resubmissions indicated that authors made substantive revisions to their submissions in response to specific feedback from the LLM. Of 40 instances when authors submitted to the Assistant multiple times, they changed their answers 39 of 40 times and often increased the length of their checklist justifications significantly. Inaccuracy (20 of 52 free-form survey responses) and excessive strictness (14 of 52) were the most frequent issues flagged by authors. For more than half of submissions, the LLM recommended changes to at least 12 of 15 checklist items. Authors most commonly reported plans to improve justifications by adding detail or citations (n = 14) and clarify experimental details or data descriptions (n = 6). On 14 of 15 questions, the Assistant could be manipulated to improve accuracy scores by changing checklist content without improving the paper.

Conclusions In this study, an LLM Checklist Assistant was shown to aid a small group of authors to ensure scientific rigor. However, the tool should not be used as a fully automated review tool that replaces human review.

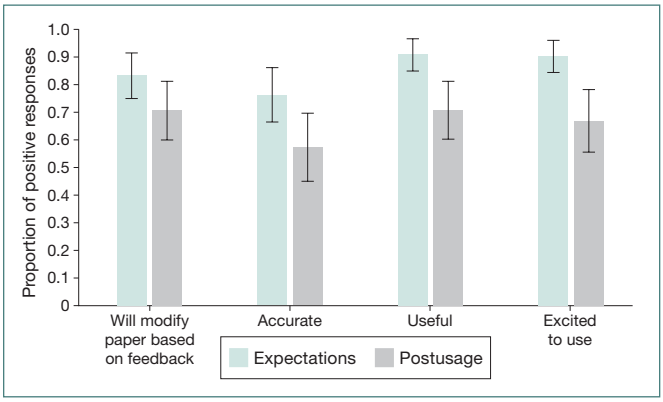
References

1. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001:1191-1194.

2. Vandenbroucke JP, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. 2014:1500-1524.

[1]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, US, akgoldbe@andrew.cmu.edu; [2]ChaLearn, Berkeley, CA, US; [3]University of Paris-Saclay, Paris, France;

Figure 25-0958. Responses to Survey Questions Pre- and Postusage of the Checklist Assistant From All Authors Who Responded to Both Surveys (N = 65)



Error bars show 95% CIs for the sample proportion. The majority of surveyed authors reported a positive experience using the Checklist Assistant.

[4]Department of Computer Science, The University of Chicago, Chicago, IL, US; [5]Google DeepMind, San Francisco, CA, US.

Conflict of Interest Disclosures Nihar B. Shah 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 research project has benefitted from the Microsoft Accelerating Foundation Models Research (AFMR) grant program, an INRIA Google Research grant, the ANR Chair of Artificial Intelligence HUMANIA ANR-19-CHIA-0022, NSF 1942124, and 2200410.

Acknowledgment In preparing this experiment, we received advice and help from many people. We are particularly grateful to the NeurIPS 2024 organizers, including General Chair Amir Globerson; Program Chairs Danielle Belgrave, Cheng Zhang, Angela Fan, Jakub Tomczak, and Ulrich Paquet; and workflow team member Babak Rahmani, for participating in brainstorming discussions and contributing to the design. We have also received input and encouragement from Andrew McCallum of OpenReview, Anurag Acharya from Google Scholar, and Tristan Neuman from the NeurIPS board. Several volunteers contributed ideas and helped with various aspects of the preparation, including Jeremiah Liu, Lisheng Sun, Paulo Henrique Couto, Michael Brenner, Neha Nayak Kennard, and Adrien Pavao. This research project has benefited from the Microsoft AFMR grant program. We are grateful to Marc Schoenauer for supporting this effort with an INRIA Google Research grant. We acknowledge the support of ChaLearn, the ANR Chair of Artificial Intelligence HUMANIA ANR-19-CHIA-0022, NSF 1942124, and 2200410. Importantly, we are thankful to all the participants of the Checklist Assistant for volunteering to try it out and providing their valuable feedback.

Accuracy and Precision of a Neural Network Author Name Disambiguator

Vicente Amado Olivo,[1] Wolfgang Kerzendorf,[1] Nutan Chen,[2] Joshua V. Shields,[1] Bangjing Lu,[1] Andreas Flörs[3]

Objective The process of identifying peer reviewers is becoming more difficult due to a surge in submissions and declining acceptance rates for review invitations.[1] Responses to a 2024 IOP Publishing online survey suggest the peer review system is unevenly distributed, with 10% of reviewers conducting 50% of all reviews, while early-career researchers and scholars from underrepresented regions remain underused despite being willing to review.[2] Reviewer identification tools have been introduced to assist editors, but the tools are proprietary and closed source, limiting their accessibility and transparency to the broader scientific community. To expand the global pool of peer reviewers, open-source, AI-powered methods are needed to uniquely identify researchers within the expanding scientific literature and match them with appropriate review opportunities. Current author name disambiguation methods often rely on extensive data features, such as institutional affiliations, email addresses, or publication venues, which are not consistently available across digital libraries. We describe the development of a name disambiguation tool with fewer data features.

Design Given the lack of existing author disambiguation tools for the Smithsonian Astrophysical Observatory/National Aeronautics and Space Administration Astrophysics Data System, we developed the Neural Author Name Disambiguator (NAND), a similarity-based neural network, and trained it on pairs of publications authored by individuals sharing the same name labeled with Open Researcher and Contributor ID (ORCID) identifiers. The training dataset included 2,698,778 pairs balanced between classes (0 if they shared the same ORCID and 1 if they had different ORCIDs), 553,496 unique publications, and 125,486 ORCID profiles of authors from the 2020 ORCID open data snapshot (and annual public data release). NAND was trained to disambiguate publications with minimal features (ie, author name, title, and abstract). We validated NAND performance using standard classification metrics: accuracy (percentage of correctly classified publication pairs), precision (true positives / [true positives + false positives]), recall (true positives / [true positives + false negatives]), and F1 scores (true positives / [true positives + 0.5(false positives + false negatives)]).

Results NAND achieved a mean (SD) 94.64% (0.04%) accuracy on test set pairs of authors within the same name block (eg, J. Smith or Y. Wang). Mean (SD) precision, recall, and F1 scores were 96.67% (0.05%), 95.21% (0.11%), and 95.94% (0.03%), respectively.

Conclusions Combining the disambiguation model with semantic expertise matching techniques could establish a practical framework for identifying qualified and willing reviewers across global institutions.[3] The open framework may help to reduce the burden on overused reviewers by expanding the global pool of available reviewers. While results are promising, this analysis was limited to physics publications; further validation is needed to assess generalizability to other scientific domains.

References

1. Publons. 2018 Global State of Peer Review. Accessed January 29, 2025. https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf

2. Brigham L, Brent-Jones E, Coombs A, et al. State of peer review 2024. Accessed January 29, 2025. https://ioppublishing.org/state-of-peer-review-2024-results/

3. Kerzendorf WE, Patat F, Bordelon D, van de Ven G, Pritchard TA. Distributed peer review enhanced with natural language processing and machine learning. *Nat Astronomy*. 2020;4:711-717. doi:10.1038/s41550-020-1038-y

[1]Michigan State University, East Lansing, MI, US, amadovic@msu.edu; [2]Volkswagen Group, Wolfsburg, Germany; [3]GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt, Germany.

Conflict of Interest Disclosures None reported.

Funding/Support This work is supported in part by the National Science Foundation Research Traineeship Program (DGE-2152014) to Vicente Amado Olivo. Additionally, we gratefully acknowledge the European Space Agency for funding support through a traineeship for Vicente Amado Olivo.

Additional Information We acknowledge the support and guidance from Markus Kissler-Patig and Jan Reerink at the European Space Agency. The authors used the free versions of ChatGPT (OpenAI) and Claude (Anthropic) for brainstorming and editing for logical flow. We take responsibility for the integrity of the content generated.

An AI-Assisted Analysis of Published *PeerJ* Open Peer Reviews

Peiling Wang,[1] Dietmar Wolfram,[2] Scott Shumate[3]

Objective Open peer review processes promote transparency and accountability. As more publishers adopt open peer review,[1] accessible peer reviews become valuable corpora to develop AI tools for enhancing peer review processes.[2] The main purpose of peer review as stated in the Guidelines for Reviewers[3] is, "First, reviews provide constructive advice and recommendations to the authors on how their paper can be improved." How can we determine the constructiveness of peer review reports? This exploratory study analyzed 11,321 peer review reports of the first version of 4100 manuscripts submitted to, and published in, *PeerJ* to identify constructive comments. Specifically, this dataset includes first reviews from November 11, 2012, to November 24, 2019. This study examines 3 research questions: (1) To what extent do reviewers provide constructive comments? (2) To what extent do reviewers use hedging to soften criticisms or recommendations? (3) Are there differences in constructive comments between signed and anonymous reviews?

Design The gpt-4o-2024-08-06 model from OpenAI API was used to analyze reviews at the sentence level to identify criticisms or supportive comments about different aspects of the research, including study purpose, hypotheses, methods, results, discussion, and conclusions. A detailed system prompt (https://zenodo.org/records/15464322) was used with a temperature setting of 0. The model was instructed to categorize each sentence into a structured JSON output specifying section, aspects, evaluation, tone, and mode. Use of hedging tone includes phrases such as "I do not think the investigation was 'rigorous'" or "has the potential to be much better and more relevant." Building on the GPT analysis, we derived the following variables to measure review feature aspects, including (1) Pc, the percentage of critical sentences; (2) Ps, the percentage of supportive sentences; and (3) Ph, the percentage of hedging sentences. The constructive index (CI), considering both criticisms and supportive comments, is the percentage of sentences containing research comments.

Results Of the 11,321 peer review reports, 7071 (62.5%) were anonymous and 4250 (37.5%) were signed (**Table 25-1055**). Few reports were supportive (overall median of 9%). Higher levels (27% of anonymous reviews and 25% of signed reviews) contained criticisms of the research. A similar level of hedging was observed. Significant differences were observed between the 2 groups; signed reviews showed significantly fewer criticisms and higher levels of hedging.

Conclusions We measured the extent of research-related comments in peer review reports that authors opted to publish. Anonymous reviewers tend to be more critical. The use of hedging in reviews could obscure clear suggestions on how to improve the manuscript. This research represents an initial effort towards a model for an AI-assisted peer review system. Further analysis should fine-tune the AI algorithms for analysis of reviewer comments. In addition, authors' perspectives on what constitutes constructive feedback warrant further study. This research had limitations. Review histories were not available for rejected manuscripts or for articles where authors opted out of publishing reviews.

Table 25-1055. Review Feature Descriptive and Nonparametrical Statistical Tests

Review feature	Median (IQR), %			Statistical significance[a] (*P*)
	All (N = 11,321)	Anonymous (n = 7071, 62.5%)	Signed (n = 4250, 37.5%)	
Criticism	26 (15-40)	27 (17-41)	25 (13-38)	<.001
Supportive	9 (0-18)	8 (0-17)	10 (3-20)	<.001
Constructive index	39 (27-53)	39 (27-53)	38 (27-51)	.03
Hedging	25 (13-38)	23 (13-36)	26 (14-39)	<.001

[a]Mann-Whitney test (skewness <1.0 except for supportive).

References

1. Wolfram D, Wang P, Hembree A, Park H. Open peer review: promoting transparency in open science. *Scientometrics*. 2020;125:1033-1051. doi:10.1007/s11192-020-03488-4

2. Wolfram D, Wang P, Abuzahra F. An exploration of referees' comments published in open peer review journals: the characteristics of review language and the association between review scrutiny and citations. *Res Eval*. 2021;30(3):314-322. doi:10.1093/reseval/rvab005

3. Technical Community on Real-Time Systems (TCRTS) of the IEEE Computer Society. Guidelines for reviewers. Accessed on May 29, 2025. https://cmte.ieee.org/tcrts/guidelines-for-reviewers

[1]University of Tennessee, Knoxville, US, peilingw@utk.edu; [2]University of Wisconsin, Milwaukee, US; [3]Austin Peay State University, Clarksville, TN, US.

Conflict of Interest Disclosures None reported.

Comparing Observational Exposure-Phenotype Correlations With Large Language Model Predictions

Chirag J. Patel,[1] Arjun K. Manrai,[1] Randall J. Ellis,[1] John P. A. Ioannidis[2]

Background Large language models (LLMs) are increasingly used to synthesize biomedical evidence, yet it is unknown whether their inferred exposure-phenotype relationships mirror observational findings.[1,2]

Objective To quantify concordance of exposure-phenotype empirical associations using National Health and Nutrition

Examination Survey (NHANES) data with predictions from an LLM (ChatGPT 4o [OpenAI]).

Methods We computed 115,056 exposure-phenotype partial correlations among 560 chemical, dietary, and clinical exposures and 268 phenotypes in 80,000 US adults (1999-2018 NHANES).[3] We then sampled 1500 pairs: 500 significant after Bonferroni correction, 500 significant only after false-discovery rate (FDR) correction, and 500 not significant—to fit within the LLM prompt window. An OpenAI-o1-aug24 model (temperature = 0) acting as a prompted "epidemiologist" was asked to predict the sign, magnitude, and *P* value of each pair and rated the evidence generated as high, trending, or borderline. Verifiability of LLM-cited references was assessed by estimating κ values and Spearman correlation.

Results The LLM labeled 23 pairs as high evidence, 244 as trending evidence, and 1233 as borderline evidence. Sign concordance with NHANES was 73% for high-confidence pairs (κ = 0.49), 65% for Bonferroni pairs (κ = 0.25), 43% for FDR-only pairs (κ = 0.09), and 12% for borderline pairs (κ approximately 0) (**Figure 25-1123**). Random sign assignment yielded a κ of 0.0001. Concordance for correlation magnitudes was modest (Spearman ρ = 0.21). Permutation tests confirmed no greater-than-chance performance on nonsense data. Mechanistic explanations dominated high-confidence pairs; 41% of cited references (615 of 1500) were unverifiable.

Conclusions LLMs captured directionality for well-established exposure-phenotype associations but faltered on borderline signals, highlighting both the promise and current limits of LLM-based evidence synthesis in exposomic epidemiology.

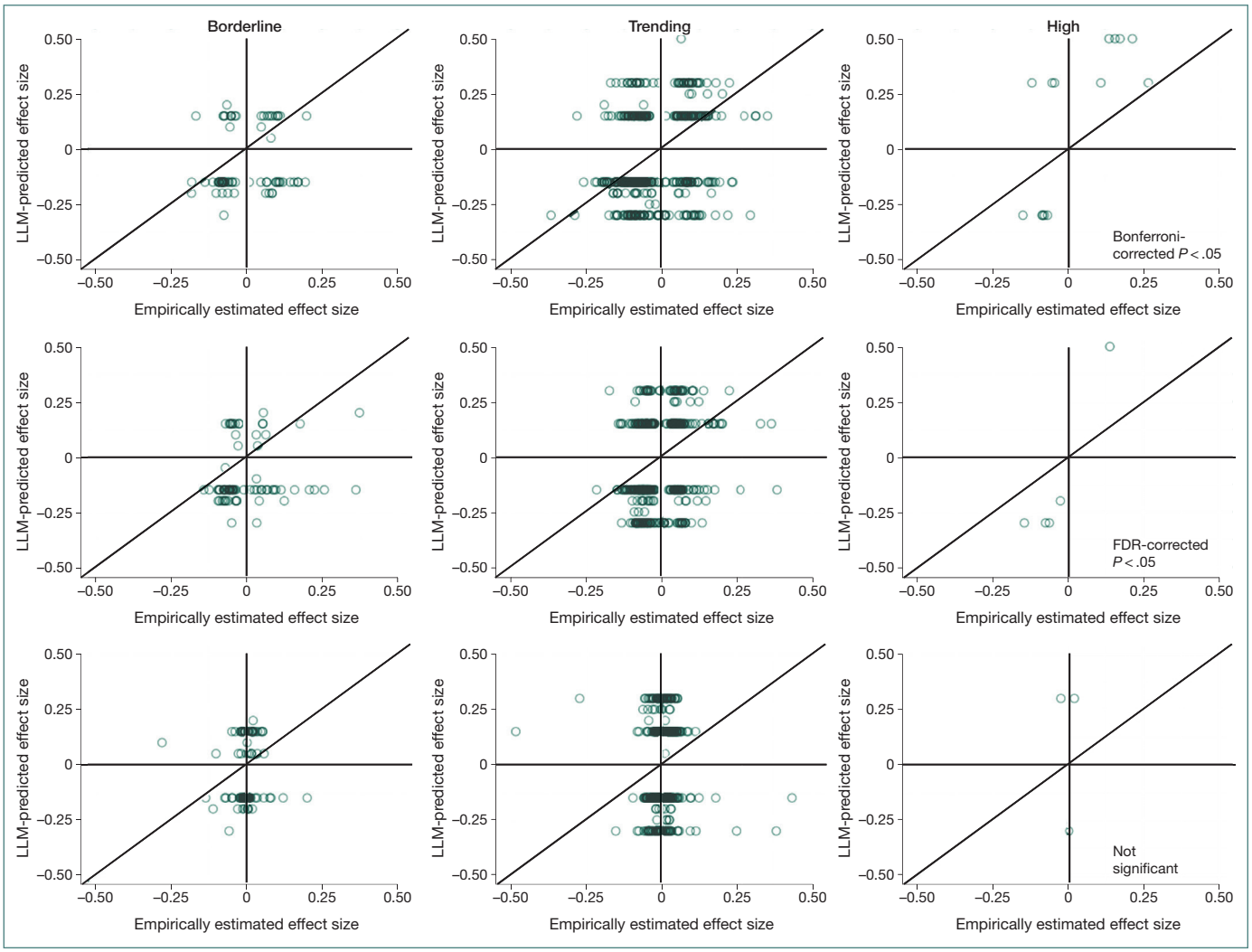
References

1. Ioannidis JPA, Loy EY, Poulton R, Chia KS. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci Transl Med*. 2009;1(7):7ps8. doi:10.1126/scitranslmed.3000247

2. Patel CJ, Ioannidis JPA. Studying the elusive environment in large scale. *JAMA*. 2014;311(21):2173-2174. doi:10.1001/jama.2014.4129

3. Patel CJ, Rehkopf DH, Leppert JT, et al. Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States

Figure 25-1123. Empirically Estimated and LLM-Predicted Effect Sizes



National Health and Nutrition Examination Survey. *Int J Epidemiol*. 2013;42(6):1795-1810. doi:10.1093/ije/dyt208

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, US, chirag@hms.harvard.edu; [2]Department of Medicine, Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, US.

Conflict of Interest Disclosures John P. A. Ioannidis 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 grants R01ES032470 and U24ES036819 from the National Institutes of Environmental Health Sciences and grant R01DK137993 from the National Institute of Diabetes and Digestive and Kidney Diseases.

Role of Funder/Sponsor The sponsors had no role in this study.

Pragmatic Assessment of Different AI Large Language Models for Extraction of CONSORT Items From Randomized Controlled Trials Before Peer Review

Nicola Di Girolamo,[1,2] Reint Meursinge Reynders,[3,4] Ugo Di Girolamo[5]

Objective Consolidated Standard of Reporting of Trials (CONSORT) checklists help authors to adhere to reporting standards and allow editors and peer reviewers to ensure that critical information is reported in a randomized controlled trial (RCT) at a glance.[1] We pragmatically assessed the ability of 3 artificial intelligence (AI) large language model (LLM) chatbots provided with minimal instructions to properly complete a CONSORT checklist from manuscripts reporting RCTs before peer review in 2025.

Design This cross-sectional study was registered on the Open Science Framework.[2] PDFs of manuscripts reporting RCTs before peer review that were consecutively published in *The BMJ* were extracted in reverse-chronological order for a total of 50 manuscripts. *The BMJ* granted permission for this study. In February 2025, each PDF was uploaded exactly as submitted by the authors to 3 LLMs (ChatGPT-4o, Gemini Advanced 2.0 Flash, and Claude 3.5 Sonnet), which were instructed to complete the CONSORT checklist in a table format,[3] including the reporting of the 37 CONSORT items and extraction of the text where the item was reported. The number of iterations and total time (seconds) required to obtain a visually acceptable table were recorded. Sensitivity, specificity, and accuracy for each LLM were calculated with human manual extraction as the reference standard. Manual extraction was performed by 1 operator (N.D.) with expertise in RCT methods. Generalized linear mixed models were built to evaluate the effect of PDF characteristics (number of pages, file size, and number of words) on the disagreements between AI and human assessment.

Results Of the 1850 CONSORT items, ChatGPT provided decisions for 1776 (96.0%); 1159 of these were in agreement with human extraction (overall accuracy, 65.3% [95% CI, 63.0%-67.5%]; sensitivity, 61.3% [95% CI, 55.0%-67.3%]; specificity, 65.9% [95% CI, 63.5%-68.3%]). Gemini provided decisions for 1629 items (88.0%), of which 1521 were in agreement with human extraction (accuracy, 93.4% [95% CI, 92.1%-94.5%]; sensitivity, 97.6% [95% CI, 94.8%-99.1%]; specificity, 92.6% [95% CI, 91.1%-94.0%]). Claude provided decisions for 1480 items (80.0%), of which 1375 were in agreement with human extraction (accuracy, 92.9% [95% CI, 91.5%- 94.2%]; sensitivity, 56.1% [95% CI, 49.3%-62.8%]; specificity, 99.4% [95% CI, 98.8%-99.7%]). Accuracy of LLMs was particularly low for certain CONSORT items (**Figure 25-1152**). There were no significant associations between PDF characteristics and AI and human disagreements. Trial PDFs were a mean of 49 pages and 17,268 words. LLMs were able to compile visually acceptable CONSORT checklists on the first iteration in 76.7% (115 of 150) attempts: ChatGPT for 40 RCTs, Gemini for 36 RCTs, and Claude for 39 RCTs. Generation of a CONSORT checklist required a mean of 75 seconds, 38 seconds, and 47 seconds for ChatGPT, Gemini, and Claude, respectively (range, 16-249 seconds).

Conclusions In early 2025, the LLMs Gemini and Claude, but not ChatGPT, were able to generate accurate CONSORT checklists from PDFs of manuscripts reporting RCTs before peer review in over 90% of cases, given minimal instructions and often in less than 1 minute.

References

1. Moher D, Jones A, Lepage L; CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*. 2001;285(15):1992-1995. doi:10.1001/jama.285.15.1992

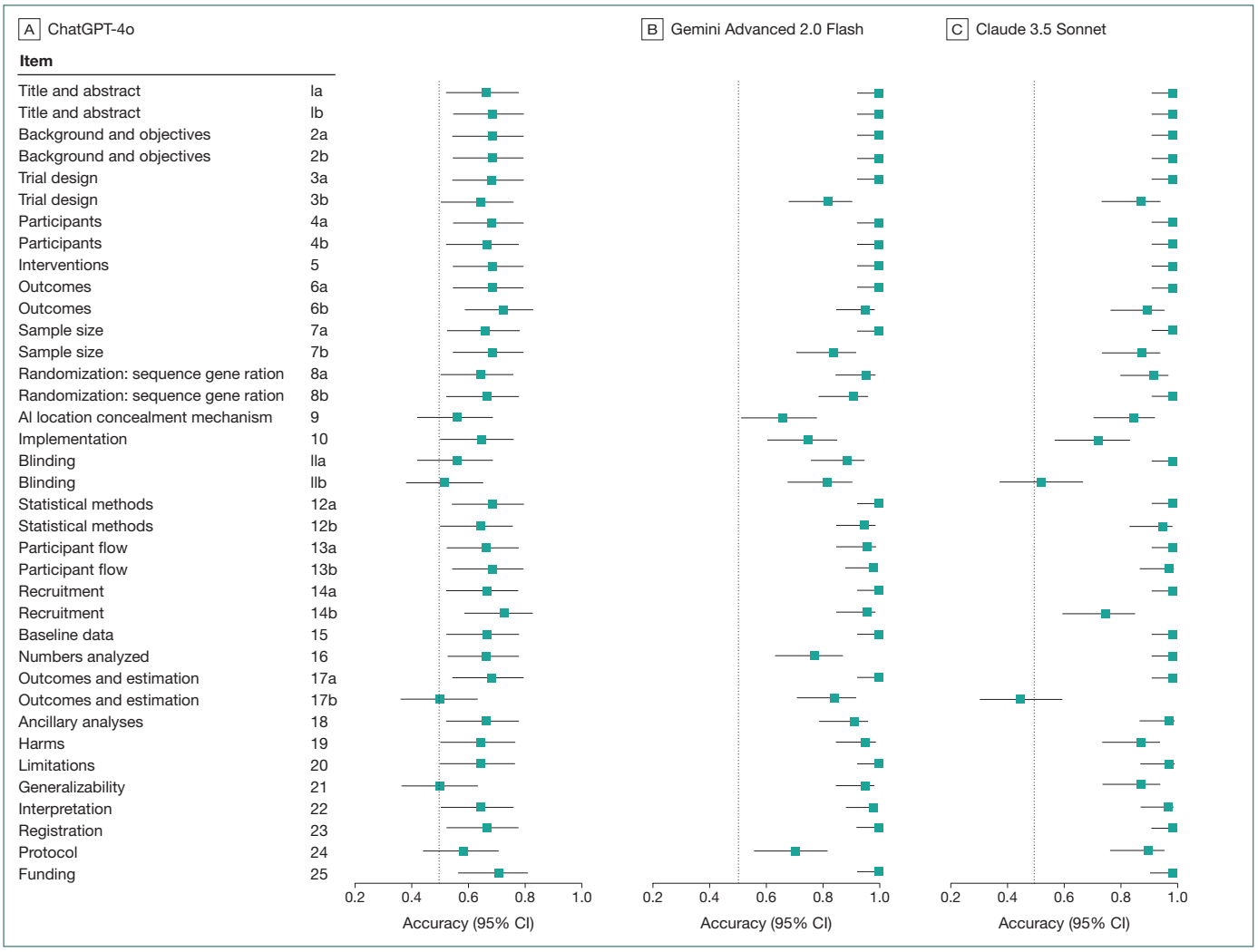
2. Di Girolamo N, Meursinge Reynders R, Di Girolamo U. Comparison of different AI large language models for extraction of CONSORT items from submitted randomized controlled trials before peer-review. Open Science Framework. Registered February 13, 2025. Accessed July 5, 2025. https://osf.io/5v6h2

3. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. doi:10.1136/bmj.c332

[1]Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, US, nd374@cornell.edu; [2]Journal of Small Animal Practice, British Small Animal Veterinary Association, Gloucestershire, UK; [3]Department of Oral and Maxillofacial Surgery, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, the Netherlands; [4]Private Practice of Orthodontics, Milan, Italy; [5]Compass, New York, NY, US.

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

Figure 25-1152. Overall Accuracy of 3 Artificial Intelligence (AI) Large Language Models (LLMs) for Extracting 37 CONSORT Items



Data are from PDFs of manuscripts reporting randomized controlled trials (RCTs) exactly as submitted by the authors to *The BMJ* and before peer review. The authors did not use extensive training or prompt engineering to assist the AI LLM or modify the PDF of the RCTs.

Artificial Intelligence Editorial Policies and Reporting Standards in Orthopedic and Sports Medicine Journals

Josh Major,[1] Kurt Mahnken,[1] Alec Young,[1] Cameron O'Brien,[1] Andrew V. Tran,[1] Patrick Crotty,[1] Alicia Ito Ford,[1,2] Matt Vassar[1,2]

Objective The integration of artificial intelligence (AI) in scientific research presents both opportunities and challenges, particularly concerning transparency, ethics, and reproducibility.[1] This study aimed to evaluate how leading orthopedic and sports medicine (OSM) journals address 2 distinct outcomes: (1) the extent to which AI is permitted in research and publication and (2) the degree to which OSM journals are endorsing AI-specific reporting guidelines (RGs). The primary outcome was to assess the presence of AI-related policies, including requirements for AI disclosure, acceptance or prohibition of AI-generated images and content, and AI's role in manuscript writing and authorship. Secondary outcomes examined the endorsement of AI-specific RGs and the relationship between AI policies and Journal Impact Factor.

Design We conducted a cross-sectional analysis of the top 100 OSM journals ranked by the 2023 SCImago Journal Rank (SJR) indicator. Eligibility criteria included actively publishing peer-reviewed clinical journals with publicly accessible Instructions for Authors in English language. On September 29, 2024, 2 investigators independently screened, extracted, and reconciled data in a masked, duplicate manner. Data were extracted from journals' Instructions for Authors pages and affiliated publisher websites where indicated. If a journal had no mention of AI policies or AI-specific RGs, the editorial teams were contacted directly to inquire about these gaps. Biserial correlation analyses in R version 4.4.1 (The R Foundation) and RStudio (Posit) examined AI policy relationships with Journal Impact Factor and SJR.

Results The initial search yielded 319 journals, from which the top 100 were analyzed. Of these, 94 mentioned AI in their Instructions for Authors, with all 94 requiring disclosure of AI use and prohibiting AI as an author (**Table 25-1175**).

Table 25-1175. Artificial Intelligence (AI) Guidelines of Orthopedic and Sports Medicine Journals

Characteristic	Journals, % (n = 94)
AI tools allowed for authorship	
Yes	0
No	100
Require authors to disclose use of AI during submission	
Yes	100
No	0
AI tools allowed in manuscript writing	
Yes	100
No	0
AI tools allowed in content generation	
Yes	87.2
No	12.8
AI tools allowed in image generation	
Yes	63.8
No	36.2

Additionally, AI-assisted manuscript writing was permitted in all cases, while 87% allowed AI-assisted content generation and 64% permitted AI-generated images. Despite this widespread recognition of AI policies, AI-specific RGs were endorsed in only 1% of journals, recommending the use of the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) RG. No significant associations were observed between Journal Impact Factor and the presence of AI-related editorial policies. Additionally, no differences were found among the AI policies in the 4 quartiles of the OSM journals in this study.

Conclusions While some AI use in research is widely acknowledged by OSM journals, the lack of consistency in policies regarding AI-generated images and content suggest an area for improvement to standardize how AI can be applied in these journals. Furthermore, the lack of endorsement for AI-specific RGs suggests a critical gap in ensuring the transparency and methodological rigor of AI-integrated research. To address these issues, we recommend that OSM journals establish clear AI policies and endorse AI-specific RGs. The International Committee of Medical Journal Editors, Committee on Publication Ethics, and World Association of Medical Editors have taken initial steps; however, implementation may be difficult.[2] Implementing these measures will bridge existing gaps, promote transparency, and improve research quality in scientific publishing.

References

1. Prasana P, Mandal PK, Hussain D, et al. Intersection of orthopaedics and artificial intelligence: a review. *SSR Inst Int J Life Sci*. 2024;10(3):5544-5552. doi:10.21276/ssr-iijls.2024.10.3.21

2. Simera I, Moher D, Hoey J, Schulz KF, Altman DG. The EQUATOR Network and reporting guidelines: helping to achieve high standards in reporting health research studies. *Maturitas*. 2009;63(1):4-6. doi:10.1016/j.maturitas.2009.03.011

[1]Office of Medical Student Research, Oklahoma State University Center for Health Sciences, Tulsa, OK, US, youngalec9.r@gmail.com; [2]Department of Psychiatry and Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, OK, US.

Conflict of Interest Disclosures Alicia Ito Ford reports receipt of funding from the Center for Integrative Research on Childhood Adversity, the Oklahoma Shared Clinical and Translational Resources, and internal grants from Oklahoma State University and Oklahoma State University Center for Health Sciences outside of the present work. Matt Vassar reports receipt of 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 outside of the present work. No other disclosures were reported.

Additional Information Andrew V. Tran is a co–corresponding author (andrewtranresearch@gmail.com).

Enhancing Research Integrity in Abstract Submissions With a Hybrid AI-Human Review Process

Heather Goodell,[1] Christine Beaty,[1] Jonathan Schultz,[1] Shilpi Mehra,[2] Chirag Jay Patel[3]

Objective The increase in abstract submissions to major scientific conferences requires efficient and reliable methods to ensure compliance with research integrity standards, which are essential for maintaining the credibility of science. Traditional review methods may not fully capture key issues, requiring a structured approach to evaluate submissions before final acceptance. This study presents a review process that uses artificial intelligence (AI) with human validation to assess the integrity of abstracts submitted to the American Heart Association (AHA) Scientific Sessions.

Design An integrity evaluation process was applied by combining AI-driven analysis with human review of abstracts submitted for the 2024 AHA Scientific Sessions. An AI tool (Paperpal Preflight for Editorial Desk) was used to identify research integrity issues, including AI-generated text, author expertise misalignment, and reference accuracy. Subject matter experts then reviewed abstracts flagged for validation. AI checks categorized abstracts into *passed*, *warning*, or *critical* levels, with those not marked as *passed* requiring further human review.

Results Among the 8477 submitted abstracts that were analyzed, 42 were flagged with a warning for integrity concerns (AI-generated text, author expertise misalignment, and reference accuracy); human reviewers cleared 3 of these. Further analysis revealed that 167 authors submitted 15 or more abstracts, 63 authors submitted 20 or more, and 13 authors submitted more than 30 abstracts. A total of 2624 abstracts had at least 1 author with 10 or more submissions,

and 1438 abstracts had at least 1 author with 15 or more submissions. The number of submissions was considered when assessing research integrity concerns. Analysis of author submission counts helped flag questionable cases for closer review and the identification of potentially suspect activities.

Conclusions The hybrid AI-human review process for abstract submissions was demonstrated to be an efficient and accurate evaluation method that should be scalable. By addressing concerns related to author submission volume and content errors, this approach can strengthen conference proceedings and support broader efforts to enhance research quality and trust in scientific dissemination. Future efforts will focus on refining and customizing AI models and optimizing reviewer workflows to further enhance this process.

[1]American Heart Association, Dallas, TX, US; [2]Cactus Communications Pvt Ltd, Mumbai, India; [3]Cactus Communications Inc, Princeton, NJ, US; chirag.patel@cactusglobal.com.

Conflict of Interest Disclosures Heather Goodell, Christine Beaty, and Jonathan Schultz are employed by the American Heart Association. Shilpi Mehra and Chirag Jay Patel are employed by Cactus Communications, which owns Paperpal Preflight for Editorial Desk. No other disclosures were reported.

Virtual

Policies on Artificial Intelligence Among Academic Publishers

Jeremy Y. Ng,[1,2,3] Daivat Bhavsar,[4] Laura Duffy,[4] Hamin Jo,[4] Cynthia Lokker,[4] R. Brian Haynes,[4,5] Alfonso Iorio,[4,5] Ana Marušić[6]

Objective This study examined the policies implemented by academic publishers regarding authors' use of generative artificial intelligence (GenAI) tools, focusing on their regulation, disclosure requirements, and role in ensuring the integrity of scientific publications. By analyzing the prevalence and content of these policies, this study aimed to provide insight into the current landscape and inform future policy development in the rapidly evolving field of artificial intelligence (AI)–assisted research and publication.

Design A cross-sectional audit was conducted on the publicly available policies of 163 academic publishers listed as members of the International Association of Scientific, Technical, and Medical Publishers. Policies were collected and analyzed between September 1 and December 31, 2023. Publishers without publicly accessible policies specific to GenAI use by authors were excluded. Data extraction and analysis were conducted independently in duplicate, with a third reviewer resolving discrepancies. The key policy components analyzed included authorship accreditation, disclosure requirements, and permissions for tasks such as research methods, content generation, image creation, and proofreading. Descriptive statistics were used to summarize the findings. Our protocol was registered.[1]

Results Of 163 academic publishers, 56 (34.4%) had publicly available policies guiding GenAI use by authors. None permitted authorship accreditation for AI tools, citing accountability concerns and alignment with ethical guidelines. Nearly all publishers with policies (49 of 56 [87.5%]) mandated disclosure of GenAI use, primarily in the Methods or Acknowledgments section. However, disclosure practices varied, with some publishers providing standardized templates while others left requirements vague. Four publishers completely prohibited GenAI use in manuscript preparation, while others allowed their use for specific tasks. Most (33 of 56 [58.9%]) publishers permitted GenAI for drafting nonmethodological sections (eg, Introductions), while 18 (32.1%) permitted their use in research methods, such as data analysis and organization. Few publishers addressed GenAI use in image generation (14 of 163 [8.6%]) or proofreading (15 of 163 [9.2%]). Only 1 publisher (0.6%) allowed citation of AI as primary sources, while 19 (11.6%) explicitly prohibited such citations. Our study has been published.[2]

Conclusions This audit highlights the inconsistent development of GenAI policies among academic publishers, with large variability in scope and clarity. While the prohibition of AI authorship and the emphasis on mandatory disclosure are consistent themes, inconsistencies in regulating specific tasks suggest a need for standardized and comprehensive policies. As AI technology and its applications in research evolve, publishers must adapt to safeguard scientific integrity. Given that the AI landscape is fast moving, future research includes updating this audit and comparing and contrasting current policies with those found in this study. Future research should also assess how policies are implemented and enforced by examining samples of published articles, as well as explore how these policies affect editors and reviewers, taking into account potential risks, such as privacy breaches and bias.

References

1. Bhavsar D, Lokker C, Haynes RB, Iorio A, Marusic A, Ng JY. Academic publisher artificial intelligence chatbot policies for authors: a cross-sectional audit. OSF Registries. Accessed July 11, 2025. doi:10.17605/OSF.IO/937ES

2. Bhavsar D, Duffy L, Jo H, et al. Policies on artificial intelligence chatbots among academic publishers: a cross-sectional audit. *Res Integr Peer Rev*. 2025;10(1):1. doi:10.1186/s41073-025-00158-y

[1]Institute of General Practice and Interprofessional Care, University Hospital Tübingen, Tübingen, Germany, jeremyyng.phd@gmail.com; [2]Robert Bosch Center for Integrative Medicine and Health, Bosch Health Campus, Stuttgart, Germany; [3]Centre for Journalology, Ottawa Hospital Research Institute, Ottawa, Canada; [4]Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; [5]Department of Medicine, McMaster University, Hamilton, Ontario, Canada; [6]Department of Research in Biomedicine and Health and Center for Evidence-Based Medicine, University of Split School of Medicine, Split, Croatia.

Conflicts of Interest Disclosures The authors declare no conflicts of interest. Ana Marušić is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Use of Generative Artificial Intelligence Tools by Authors and Reviewers of the Journal *Eurosurveillance*

Eva Sarachaga,[1] Ines Steffens[1]

Objective In September 2023, *Eurosurveillance* adopted an artificial intelligence (AI) policy encouraging authors and reviewers to use AI tools responsibly. The policy states that authors should disclose tools such as large language models (LLMs), chatbots, and image-generating algorithms used in the production and writing of the manuscript. It further outlines that reviewers should declare whether they have used a chatbot or LLM tool in the generation of their review or in their correspondence to authors or editors.[1] Fifteen months after launching the policy, we investigated how many authors and reviewers had declared using AI tools and how. Additionally, we examined whether possible AI use was flagged by the plagiarism detection system, iThenticate, which checks whether incoming submissions overlap with published material.

Design Data were extracted from the article submission system, Editorial Manager. In a cross-sectional design, we included all articles and reviews in the system since the policy was published in September 2023 until November 2024. Results were analyzed by domain, AI tool, and tool version. To investigate whether results from iThenticate differed between articles where AI use was declared versus where it was not declared, we compared scores of a convenience sample of 43 articles in each group and checked manually thereafter.

Results Only 7% of authors (70 of 961) and 2% of reviewers (11 of 707) declared using AI tools. Language was the most frequent domain in both groups: 61 of 70 authors and 10 of 11 reviewers. Authors declared other domains, such as coding, data analysis, prediction identification, correspondence, and figure creation. ChatGPT was the most used AI tool in both groups: 54 of 70 authors and 6 of 11 reviewers. The second most used AI tools were DeepL for authors (2 of 70) and Paperpal for reviewers (2 of 11). Our results were in line with results from other studies.[2,3] AI use was not declared according to the policy: only 20% of authors (13 of 70) and 20% of reviewers (2 of 11) declared the tool version as required. The plagiarism score was lower in articles where AI use was declared (35 of 43) versus where it was not declared (39 of 43).

Conclusions In line with findings by others, only a limited proportion of authors and reviewers declared using AI tools following implementation of an AI policy. The proportion was similar for authors and reviewers, with language being the most frequent domain and ChatGPT the most used tool in both groups. Authors and reviewers need to be reminded to comply better with the AI policy and to declare tool version.

References

1. E*urosurveillance*. Editorial policy: Responsible use of artificial intelligence (AI) tools. Accessed July 10, 2025. https://www.eurosurveillance.org/editorial-policy#AI%20policy

2. Salvagno M, De Cassai A, Zorzi S, et al. The state of artificial intelligence in medical research: a survey of corresponding authors from top medical journals. *PLOS One*. 2024;19(8):e0309208. doi:10.1371/journal.pone.0309208

3. Else H. Should researchers use AI to write papers? Group aims for community-driven standards. *Science*. 2024;384(6693). doi:10.1126/science.z9gp5z0

[1]European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden, evasarachaga01@gmail.com.

Conflict of Interest Disclosures None reported.

Acknowledgment We would like to thank the *Eurosurveillance* editorial team (Alina Buzdugan, Anirban Dey, Kathrin Hagmaier, Megan Osler, Elina Tast-Lahti, and Karen Wilson), with special thanks to Kathrin Hagmaier, at the ECDC who provided their feedback on this work.

Use of an AI Peer Review Panel to Assess Manuscript Clarity, Novelty, and Impact

Pawin Taechoyotin,[1] Daniel E. Acuna[1]

Objective Nowadays, there is an excessive load on human reviewers to the point where the quality of peer reviews is often compromised. To alleviate this excessive load, we explored the possibility of a multiagent AI peer review panel that we have developed that considers text, figures, and citations to produce peer reviews. This study is based on the advancement of large language models[1,2] and the study of AI agents for peer review,[3] which so far only utilize the textual content within the manuscript.

Design The AI panel consisted of a leader agent, an experiment agent, an impact agent, and a clarity agent. Each agent utilized zero-shot prompting to the LLM model Claude Sonnet 3.5. Prompt engineering was used to craft the system and task prompt. The system prompts guided the agents to assume their specific role, and the task prompts provided clear instructions on the tasks to perform. Novelty was assessed by comparing the manuscript with similar articles from an external database of published literature. The manuscript, novelty assessment, and figures were sent to each agent for analysis, and the leader agent compiled the information from all agents, requested clarifications if needed, and produced the final review. Our AI multiagent panel was evaluated and compared with human reviews as well as simple prompt-based review[2] and findings of 2 previous studies.[1,3] Evaluation was done by graduate students via an arena where pairs of systems were selected based on their Elo score. The Elo score is a measure of the relative capability of each system based on their win/loss in the arena against other systems and was calculated using the Bradley-Terry model. The system names were hidden from the

Table 25-0971. Elo Score or Relative Capability of Each Model

Model	Mean ± SD				
	Technical quality	Construc-tiveness	Clarity	Overall quality	Style-adjusted score
Human reviewer	1236 ± 134	1257 ± 71	1237 ± 127	1229 ± 165	1188 ± 567
Prompt-based[2]	1391 ± 128	1392 ± 58	1387 ± 133	1396 ± 160	1396 ± 542
Liang et al,[1] 2024	1237 ± 125	1251 ± 49	1242 ± 123	1217 ± 157	1227 ± 544
D'Arcy et al,[3] 2024	1766 ± 180	1731 ± 60	1761 ± 181	1769 ± 231	1776 ± 816
Our system	1870 ± 184	1869 ± 67	1874 ± 181	1889 ± 232	1914 ± 820

participants to reduce bias. The participants evaluated the reviews based on technical quality, constructiveness, clarity, and overall assessment before making their choice.

Results Eleven graduate students in the computer science department participated in this study, which resulted in 140 matches between pairs of systems. The systems with similar Elo scores were most likely to be paired together. Mean Elo scores of the reviews produced from our system were higher than those of human review scores (**Table 25-0971**). Further analysis showed that reviews produced by our system included more detailed assessment of the limitations and possible improvements of the manuscript than reviews produced by humans. However, our system tended to overly praise work, which humans rarely do.

Conclusions Our AI peer review system demonstrated potential in aiding the peer review process of publications as an addition to human reviews. It might act as a quick response to authors while human reviewers are reviewing the manuscript. Although there might be bias within the review, the author can choose to verify and utilize comments from this quick feedback as needed. Harm or the potential for negative effects were not studied. We acknowledge that the number of participants in this study was small. Thus, a larger study might have different results from our findings.

References

1. Liang W, Zhang Y, Cao H, et al. Can large language models provide useful feedback on research papers—a large-scale empirical analysis. *NEJM AI*. 2024;1(8):AIoa2400196. doi:10.1056/AIoa2400196

2. Wang H, Fu T, Du Y, et al. Scientific discovery in the age of artificial intelligence. *Nature*. 2023;620(7972):47-60. doi:10.1038/s41586-023-06221-2

3. D'Arcy M, Hope T, Birnbaum L, Downey D. Marg: multiagent review generation for scientific papers. *arXiv*. Preprint posted online January 8, 2024. doi:10.48550/arXiv.2401.04259

[1]Science of Science and Computational Discovery Lab, Department of Computer Science, University of Colorado Boulder, Boulder, CO, US, daniel.acuna@colorado.edu.

Conflict of Interest Disclosures Daniel E. Acuna reported being the founder of ReviewerZero AI LLC (https://www.reviewerzero.ai), a company focused on research integrity solutions, which may be relevant to the topic of this publication.

Reviewer Rating Variability and Confidence and Language Model Sentiment Prediction of Machine Learning Conference Papers

Yidan Sun,[1] Mayank Kejriwal[2]

Objective Platforms like OpenReview[1] have become a popular standard for open peer review in conference-oriented fields such as computing.[2] Using OpenReview data from a machine learning conference, we examined (1) whether higher reviewer confidence is associated with more polarized scoring and (2) whether greater score variability is associated with rejection among papers with neutral mean scores. We also explored the assistive potential of a language model to assign sentiment labels to reviews of neutrally rated papers based on text alone.

Design We used the OpenReview application planning interface to collect data on all International Conference on Learning Representations 2019 Conference Track submissions submitted by the September 27, 2018, deadline. The dataset included 4332 reviews across 1419 papers. Each review contained an integer score (1-10), a confidence rating (1-5), and a final accept or reject decision from a meta-reviewer. Analyses were conducted in January 2025. We used all 4332 reviews to assess whether reviewer confidence was associated with score extremity, defined as the absolute difference between a review score and the rubric midpoint (5), using Pearson correlation. We focused on 775 neutral papers—those with all review scores between 4 and 7 (inclusive)—and measured score variability as the SD of scores within each paper. We compared variability between accepted and rejected papers using 2-sided *t* tests. We then trained a Bidirectional Encoder Representations From Transformers (BERT)–based sentiment classifier[3] using polarized reviews scored less than 4 (negative) and greater than 7 (positive). The model achieved 90.9% accuracy on held-out review-level validation data. We then applied the model to all 2356 reviews from the 775 neutral papers and took the mean of predicted sentiment scores across each paper to produce a paper-level sentiment score. We labeled the top 246 papers—matching the number of actually accepted neutral papers—as "Accept" based on model sentiment. As a baseline, we ranked papers by mean score and labeled the top 246 as "Accept" by score.

Results Reviewer confidence was positively correlated with score extremity (r = 0.15 [95% CI, 0.10-0.20]; P < .001). Among the 775 neutral papers, 246 (31.7%) were accepted and 529 (68.3%) were rejected (**Table 25-1167**). Rejected neutral papers showed greater score variability than accepted papers (mean SD of 0.80 vs 0.59; difference, 0.21 [95% CI, 0.15-0.28]; P < .001), supporting both hypotheses. Of the 246 papers labeled "Accept" by BERT, 137 were accepted. Of the

Table 25-1167. Meta-Review Outcomes for Neutral Papers

Labelling group	Meta-review: accepted	Meta-review: rejected	Total
BERT labeled accept	137	109	246
Baseline labeled accept	203	43	246
Both BERT and baseline labeled accept	120	20	140
Both BERT and baseline labeled reject	18	388	406
Total	246	529	775

Abbreviation: BERT, Bidirectional Encoder Representations From Transformers.

246 labeled "Accept" by the baseline, 203 were accepted. Among rejected papers, both methods labeled 20 as "Accept."

Conclusions Higher reviewer confidence was associated with more decisive scoring. Rejected neutral papers showed greater disagreement among reviewers. Despite moderate scores, review text often conveyed clear positive or negative sentiment, as inferred by the BERT model.

References

1. OpenReview: an open platform for peer review. Accessed July 15, 2025. https://openreview.net

2. Ford E. Defining and characterizing open peer review: a review of the literature. *J Scholarly Pub*. 2013;44(4):311-326. doi:10.3138/jsp.44-4-00

3. Wolf T, Debut L, Sanh V, et al. Transformers: state-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics; 2020:38-45.

[1]Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, US, yidans@usc.edu; [2]Research Associate Professor, Principal Scientist, Information Sciences Institute, University of Southern California, Los Angeles, CA, US.

Conflict of Interest Disclosures None reported.

Authorship and Contributorship

In-person

Co–First Authors and Co–Corresponding Authors in the *Chinese Medical Journal* and *JAMA*

Ting Gao,[1] Xiuyuan Hao[2]

Objective Authorship confers credit and has important academic, social, and financial implications.[1] Although certain journals, such as *JAMA*, enforce strict authorship policies, the recognition of equal contributions by 2 or more co–first and/or co–corresponding authors is increasingly common in biomedical publishing.[2,3] Both the *Chinese Medical Journal* (*CMJ*) and *JAMA* are peer-reviewed general medical journals covering all major medical disciplines. This study evaluated the prevalence of co–first and co–corresponding authors in *CMJ* and *JAMA*.

Design From January 1 to January 21, 2025, we examined all articles published in *CMJ* and *JAMA* in 2024, excluding News, Editors' Notes, Corrigendum/Correction, Humanities, and miscellaneous sections, to assess the frequency of co–first and co–corresponding authorship. Other characteristics examined included article type, subject, and geographic location of corresponding authors. A multivariate logistic regression model was used to identify characteristics associated with co–first or co–corresponding authors.

Results Of the 473 articles published in *CMJ* in 2024, 233 (49.3%) included co–first authors and 180 (38.1%) included co–corresponding authors. *JAMA* published 1011 articles in 2024; among them, 42 (4.2%) had co–first authors and 18 (1.8%) had co–corresponding authors. Original articles most frequently featured co–first and co–corresponding authors in both journals (**Table 25-0882**). In *CMJ*, gastroenterology was the most frequent subject associated with both forms of shared authorship. In *JAMA*, public health had the highest number of co–first authors, while cardiology had the most co–first and co–corresponding authors. In *CMJ*, most corresponding authors of articles with co–first or co–corresponding authorship were based in China. Logistic regression analysis revealed that article type, subject, and corresponding author's geographic location were not significantly associated with shared authorship status in *CMJ*. By contrast, in *JAMA*, article type was significantly associated with both co–first and co–corresponding authorship, and corresponding author's geographic location was significantly associated with co–corresponding authorship. *CMJ* showed a higher prevalence of co–first ($\chi^2 = 431.3$; $P < .001$) and co–corresponding ($\chi^2 = 363.6$; $P < .001$) authorship than *JAMA*.

Conclusions This study demonstrates a significant difference in the prevalence of co–first and co–corresponding authorship between *CMJ* and *JAMA*. China and the United States are among the leading countries in scientific output. Compared with US-based authors, Chinese authors were more likely to designate co–first and co–corresponding authors. This trend may reflect the academic, social, and economic significance of shared authorship roles within the Chinese research evaluation system. While journals such as *JAMA* impose restrictions on shared authorship, increasing multidisciplinary collaborations and the rise of team science have led others, such as *CMJ*, to adopt more flexible policies. Establishing a standardized framework for managing co–first and co–corresponding authorship is needed.

References

1. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated May 2023. Accessed January 18, 2025. https://www.icmje.org/recommendations/

2. Aakhus E, Mitra N, Lautenbach E, Joffe S. Gender and byline placement of co-first authors in clinical and basic science journals with high impact factors. *JAMA*. 2018;319(6): 610-611. doi:10.1001/jama.2017.18672

Table 25-0882. Characteristics Associated With Co–First or Co–Corresponding Authors in the *Chinese Medical Journal* and *JAMA*

	Chinese Medical Journal (n = 473)						*JAMA* (n = 1011)					
	Co–first authors (n = 233)			Co–corresponding authors (n = 180)			Co–first authors (n = 42)			Co–corresponding authors (n = 18)		
Variable	No. (%)	OR (95% CI)	*P* value	No. (%)	OR (95% CI)	*P* value	No. (%)	OR (95% CI)	*P* value	No. (%)	OR (95% CI)	*P* value
Article type		0.89 (0.77-1.04)	.14		0.89 (0.76-1.03)	.12		0.35 (0.27-0.47)	<.001		0.35 (0.23-0.54)	<.001
Original article	85 (36.5)			59 (32.8)			29 (69.0)			12 (66.7)		
Review	34 (14.6)			30 (16.7)			1 (2.4)			1 (5.6)		
Subject		0.98 (0.89-1.09)	.77		1.03 (0.93-1.16)	.54		0.95 (0.81-1.11)	.51		1.12 (0.93-1.36)	.23
Gastroenterology	21 (9.0)			15 (8.3)			3 (7.1)			1 (5.6)		
Public health	12 (5.2)			10 (5.6)			5 (11.9)			0		
Cardiology	14 (6.0)			12 (6.7)			3 (7.1)			4 (22.2)		
Geographic location[a]		0.19 (0.02-1.64)	.13		0.03 (0.04-2.63)	.28		1.08 (0.83-1.39)	.58		1.41 (1.07-1.85)	.01
China	226 (97.0)			179 (99.4)			19 (45.2)			4 (22.2)		
United States	5 (2.1)			0			6 (14.3)			11 (61.1)		

Abbreviation: OR, odds ratio.

[a]Geographic locations of the corresponding authors.

3. Hosseini M. Equal co-authorship practices: Review and recommendations. *Sci Eng Ethics*. 2020;26(3):1133-1148. doi:10.1007/s11948-020-00183-8

[1]Editorial Department, *Chinese Medical Journal*, Chinese Medical Association Publishing House, Beijing, China; [2]Editorial Department, *Chinese Medical Journal*, Chinese Medical Association Publishing House, Beijing, China, haoxiuyuan@163.com.

Conflict of Interest Disclosures None reported.

Authorship and Contributorship Criteria and Practices at the *Annals of African Surgery*

Cecilia Munguti,[1] James Kiilu,[1] James Kigera,[1] Michael Mwachiro[1]

Objective The International Committee of Medical Journal Editors (ICMJE) authorship criteria ensure appropriate credit for significant research contributions, while the Contributor Role Taxonomy (CRediT) framework offers a detailed account of individual scholarly contributions. Despite these frameworks, honorary authorship remains widespread in health sciences,[1] and surveys from Africa have documented its prevalence and the difficulty of applying authorship standards.[2] This study evaluated the fulfillment of ICMJE authorship criteria and CRediT contributions in manuscripts submitted between 2017 and 2023.

Design This retrospective study analyzed the fulfillment of ICMJE authorship criteria and author-reported CRediT contributor roles for manuscripts submitted to *Annals of African Surgery*, a journal that endorses the ICMJE authorship criteria, between 2017 and 2023. The evaluated ICMJE criteria included criteria 1 (intellectual input) and 2 (manuscript drafting or critical revision) because these are the only author-level items captured as discrete checkboxes in the submission system. Criteria 3 and 4 were not assessed because they are captured as a collective attestation in a signed mandatory licence agreement required at submission. Data were retrieved from the journal's online manuscript management system. CRediT roles were scored using a weighted system: 3 points for lead, 2 for equal, and 1 for supporting contributions across 14 roles, with a maximum possible score of 42 converted to a percentage. Descriptive statistics summarized ICMJE compliance and CRediT role distribution.

Results A total of 1619 authors and 448 manuscripts with complete authorship data were evaluated. Among the authors, 1027 (63.4%) met both ICMJE criteria 1 and 2, 460 (28.4%) met only 1 criterion, and 132 (8.2%) met none. Of the 1619 authors, 709 (43.8%) were supervisors; among these supervisor authors, 523 (73.8%) met both criteria, 124 (17.5%) met only 1 criterion, and 62 (8.7%) failed to meet any. A total of 9610 CRediT contributor roles were analyzed. On weighted CRediT scores, 275 (17%) authors had a contribution score above 50%, while 971 (60%) scored 25% or less. Among 709 supervisors, 241 (34%) had scores above 50% and 269 (38%) scored 25% or less. Funding acquisition was credited to 296 authors (18.3%). Among the supervisor authors, 248 (35.0%) reported funding acquisition, accounting for 83.8% of all authors who claimed that role.

Conclusions Using ICMJE criteria, supervisor involvement appeared to be limited, suggesting a high prevalence of guest authorship. However, weighted CRediT scores indicated that supervisors often contributed significantly, more so than other coauthors, highlighting a disconnect between traditional authorship criteria and actual contributions. Although limited to 1 journal, this study demonstrates the value of integrating authorship criteria audits with CRediT data, while noting the potential for checkbox behavior by submitting authors who may not grasp the significance of their declarations.

References
1. Meursinge Reynders RA, ter Riet G, Di Girolamo N, Cavagnetto D, Malički M. Honorary authorship is highly prevalent in health sciences: systematic review and meta-analysis of surveys. *Sci Rep*. 2024;14:4385. doi:10.1038/s41598-024-31526-2

2. Breet E, Botha J, Horn L, Swartz L. Academic and scientific authorship practices: a survey among South African researchers. *J Empir Res Hum Res Ethics*. 2018;13(5):412-420. doi:10.1177/1556264618789253

[1]*Annals of African Surgery*, Nairobi, Kenya, ceciliamunguti@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.

Additional Information James Kiilu is a co–corresponding author (jameskiilu@annalsofafricansurgery.com).

Integration of Credit and Accountability Principles in Authorship Policies of Science Journals and Research Institutions

Mohammad Hosseini,[1] Sofie Elaine Adams,[2] Yensi Flores,[3] Kathleen Hall Jamieson,[2] Joerg Heber,[4] Jennifer Heimberg,[5] Véronique Kiermer,[6] Arthur Lupia,[7] Ana Marušić,[8] Beau Nielsen,[5] Magdalena Skipper,[9] Geeta Swamy,[10] Susan Wolf[11]

Objective Although authorship is a cornerstone of the reward system of science, it is often fraught with disagreements and questionable practices.[1] A working group (including authors of this abstract and additional experts) was convened in January 2024 by the National Academy of Sciences (NAS) Strategic Council on Research Excellence, Integrity, and Trust to examine the impact of previous recommendations on institutional and journal authorship policies.[1,2]

Design Our analysis consisted of 2 parts completed in 2024 and 2025. First, we examined authorship policies of the AAAS, AGU, APS, *BMJ*, *Cell*, ICMJE, IEEE, JACS, *Nature*, *NEJM*, PLOS, *PNAS*, and SAGE (represented in the group members who authored the NAS 2018 recommendation[2]) to examine the extent to which they had adapted policies. This review captured trends regarding authorship guidelines, use of ORCIDs, and the CRediT taxonomy. Second, we searched the websites of all R1 (very high research activity) institutions (search terms *authorship*, *authorship guidelines*, *research guidelines*, and *research integrity*) in the US (n = 146) to determine whether credit and accountability or responsibility were mentioned in authorship guidelines and, if so, whether they were cast as related concepts. We then checked whether guidelines included links to external instructions (eg, ICMJE) and encouraged meetings to discuss authorship. Both phases were followed by brainstorming sessions among working group members.

Results Of the examined journal groups, 5 adopted ICMJE guidelines, 2 referenced the NAS 2018 recommendations, and others had different recommendations. Seven groups recommended using the CRediT taxonomy, but only 3 required ORCIDs for all authors. Of the 146 R1 institutions, 59 included a credit-accountability link in their guidelines, but in only 38 cases did the credit-accountability link appear at the top of the guideline. Guidelines of 60 institutions suggested authors meet proactively to discuss authorship.

Conclusions Anchoring authorship in credit (give credit where it is due) and accountability (authors are accountable for the integrity of the work) principles[2] readily explains why questionable practices (eg, gifts, ghost writing, and coercive authorship) are problematic and helps determine when authorship credit is warranted. Providing a clear description of contributions reinforces the application of these principles and fosters transparency as an important complementary principle.[3] While journals should choose authorship guidelines that suit their context, encouraging the use of ORCID and taxonomies such as CRediT are beneficial. Aligning institutional guidelines with the credit-accountability framework offers an opportunity to improve authorship practices. We plan to engage with institutional leaders to explore ways to support these improvements (eg, by offering educational materials or courses, providing mechanisms to resolve conflicts, and offering a safe environment for discussing authorship).

References
1. National Academies of Sciences, Engineering, and Medicine. *Fostering Integrity in Research*. 2017; National Academies Press. doi:10.17226/21896

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

3. Allen L, Scott J, Brand A, Hlava M, Altman M. Publishing: credit where credit is due. *Nature*. 2014;508(7496):312-313. doi:10.1038/508312a

[1]Northwestern University, US, mohammad.hosseini@northwestern.edu; [2]University of Pennsylvania, US; [3]University College Cork, Ireland; [4]Lawrence Berkeley National Laboratory, US; [5]National Academy of Sciences, US; [6]PLOS, US; [7]University of Michigan, US; [8]University of Split, Croatia; [9]*Nature*, UK; [10]Duke University, US; [11]University of Minnesota, US.

Conflict of Interest Disclosures The authors declare no conflicting interests. Véronique Kiermer and Ana Marušić are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract.

Funding/Support This work is funded by the National Academy of Sciences (NAS). Mohammad Hosseini is funded by the National Institutes of Health's (NIH) National Center for Advancing Translational Sciences (UM1TR005121).

Role of the Funder/Sponsor The funders have not played a role in the design, analysis, decision to publish, or preparation of the manuscript, and views expressed here do not represent the views and opinions of the NAS, NIH, or US government.

Authors Who Publish in a Journal and Likelihood to Serve as Reviewers

Stephan D. Fihn,[1,3] Roy H. Perlis,[2,3] Jacob Kendall-Taylor,[3] Annette Flanagin[3]

Objective The major challenges for editors of journals include attracting submissions of high-quality manuscripts and recruiting reviewers.[1] In light of a dearth of relevant research, we sought to ascertain the association between publishing manuscripts as an author and participation in peer review for the same journal.

Design We conducted a cross-sectional study using data extracted about authors and reviewers for *JAMA Network Open*, a high-volume open access general medical journal, for the years 2019 to 2023, including number of articles submitted and published as first or corresponding author during this period, numbers of requests to review and response (submitted, declined, no response), and quality of review submitted as determined by reviewing editor on a 5-point Likert scale (excellent = 1 to poor = 5). Data were analyzed using R 4.4.2.

Results Among 38,683 invited reviewers, 30,082 (77.8%) submitted no manuscripts; 4014 (10.6%) submitted ≥1 article as first or corresponding author but had no published articles; and 4497 (11.6%) submitted ≥1 article as first or corresponding author and had ≥1 publication. **Table 25-1097** summarizes review requests to these individuals and compares their rates of accepting reviews and ratings among those who returned reviews. The mean (SD) number of invitations was greatest among published reviewers (5.1 [9.5]), followed by those with submissions but no publications (3.4 [6.1]), and then individuals with no submissions (2.7 [5.9]) ($P < .007$ by analysis of variance). Reviewers with publications were significantly more likely to accept invitations and significantly less likely to ignore invitations compared with the other 2 groups ($P < .01$ by analysis of variance and post hoc pairwise test). The absolute difference in proportion of accepted invitations was 26.2% (95% CI, 24.8%-27.6%; $P < .001$) between published and never-submitted reviewers and 11.6% (95% CI, 9.7%-13.5%; $P < .001$) between published and nonpublished submitters. Among those who returned reviews, quality (mean rating) was significantly higher among reviewers with publications (absolute difference from nonsubmitters: 0.25; 95% CI, −0.29 to −0.20; $P < .001$). Differences between submitters without publications (mean difference: 0.03) and never-submitters (mean difference: 2.5%) were minimal for both rating metrics ($P > .50$).

Conclusions Authors who published in *JAMA Network Open* were significantly more likely to accept an invitation to review even though they received nearly twice as many requests as those who did not publish. Moreover, the quality of their reviews was significantly greater. Individuals who submitted manuscripts but did not publish were intermediate between these 2 groups in terms of accepting invitations to reviews and quality of reviews. These results suggest that investigators may be willing to commit more of their academic effort to journals that publish their work. Limitations of this analysis include the inability to determine the relative timing of publication versus review; restriction of the analysis to a single high-volume journal; exclusion of co-authors; and limitation to a 4-year time frame.

Reference

1. Zupanc GKH. "It is becoming increasingly difficult to find reviewers"—myths and facts about peer review. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*. 2024;210(1):1-5. doi:10.1007/s00359-023-01642-w

[1]Department of Medicine, University Washington, Seattle, WA, US; steve.fihn@jamanetwork.org; [2]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, US; [3]JAMA Network, Chicago, IL, US.

Conflict of Interest Disclosures Stephan D. Fihn (SDF) and Roy H. Perlis (RHP) are paid as consultants for work as editors of *JAMA Network Open* (SDF and RHP) and *JAMA+ AI* (RHP). Jacob Kendall-Taylor and Annette Flanagin are on the editorial staff of the JAMA Network. Annette Flanagin is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract. None of the authors have any other conflicts of interest to disclose.

Funding/Support The authors received no direct funding for this work apart from that described above.

Table 25-1097. Review Activity in the Context of Whether Reviewer Had Submitted or Published One or More Manuscripts in *JAMA Network Open*

		Reviews requested		Reviews completed		Mean proportion, %			Review quality	
Manuscript submission status[a]	No. (%)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Accepted	Declined	Ignored	Mean (SD)	Mean proportion very good/ excellent, %
Never submitted	30,082 (77.8)	2.7 (5.9)	1 (1-2)	0.5 (3.1)	0 (0-1)	19	36	0.44	1.8 (1)	80
Submitted, none accepted	4104 (10.6)	3.4 (6.1)	2 (1-3)	1 (3.2)	0 (0-1)	34	35	0.31	1.8 (0.9)	79
Submitted and accepted	4497 (11.6)	5.1 (9.5)	2 (1-5)	2 (5.8)	1 (0-2)	45	34	0.2	1.5 (0.7)	88

[a]As first or corresponding author.

Virtual

Compiling the Publications Produced by Medical Writing

Maud Bernisson[1,2]

Objective Researchers have raised concerns about scientific articles written by medical education and communication companies (MECCs) after litigations revealed that MECCs have been using unethical practices like ghostwriting.[1] Little is known about the literature produced by medical writing. While previous methods to assess this literature include surveying authors about ghostwriting,[2,3] this study offers a new method to collect this literature and makes it available to further explore medical writing.

Design The method of collection of this literature used 3 different sources: (1) Web of Science, from which I collected 26,858 DOIs of articles, abstracts, and documents mentioning "medical writ*" in the acknowledgments; (2) the websites of 15 MECCs, found during the mapping of MECCs for another study, that publicize the literature they produce, and from which I collected 4555 additional DOIs; and (3) the Industry Documents Library of University of California, San Francisco (https://www.industrydocuments.ucsf.edu), in which I found the publication trackers of 5 companies and collected 229 DOIs of publications. The final database includes the metadata of 31,642 documents published between 1972 and 2025, collected from Web of Science, cleaned in R, and analyzed with the library dplyr (version 1.1.4).

Results Transparency practices of disclosure of medical writing differed according to data sources. Only 252 of 4784 publications (5.3%) acknowledged medical writing in the Industry Documents Library and the MECCs websites. In contrast, a timeline of the literature collected from Web of Science identified an increase in transparency practices, with 138 publications acknowledging medical writing in 2008, 1475 in 2016, and 4265 in 2024. Topics also differed according to data sources. Four of the top 10 publishing journals of the MECC literature focused on health economics and outcomes research (*Value in Health*, *Journal of Medical Economics*, *Current Medical Research and Opinion*, *Pharmacoeconomics*). These 4 journals published 348 documents written by MECCs, which represents 7.6% of the data collected from MECCs websites. In this study, journals that published the MECC-written literature covered a range of topics. Among the journals assessed, a rheumatology journal published 605 MECC-written publications from 2008 to 2024; of these, 524 (86.6%) were meeting abstracts, which represented 43.3% of the 1211 meeting abstracts in the database. Limitations of this study include the difficulty in accessing sources for collecting this literature and that medical writing is not always acknowledged in published articles.

Conclusions This study highlights the need to further explore use and disclosure of medical writing in the health economics and outcomes research literature, as it targets decision-makers such as payers and policymakers. The database used in this study offers different routes to study medical writing in more detail.

References

1. Sismondo S. *Ghost-Managed Medicine: Big Pharma's Invisible Hands*. Mattering Press; 2018.

2. Flanagin A, Carey LA, Fontanarosa PB, et al. Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *JAMA*. 1998;280(3):222-224. doi:10.1001/jama.280.3.222

3. Moffatt B, Elliott C. Ghost marketing: pharmaceutical companies and ghostwritten journal articles. *Perspect Biology Med*. 2007;50(1):18-31. doi:10.1353/pbm.2007.0009

[1]LISIS, CNRS, INRAE, Université Gustave Eiffel, France, maud.bernisson@cnrs.fr; [2]Institute for Science in Society, Radboud University, the Netherlands.

Conflict of Interest Disclosures None reported.

Funding/Support The author was part of the NanoBubbles project, which has received Synergy grant funding from the European Research Council, within the European Union's H2020 programme (grant agreement 951393).

Additional Information The database for this study, including the Web of Science identifiers, DOIs (when they exist), and sources (MECCs websites, Industry Documents Library, and Web of Science) of 31,642 documents involving medical writing, collected on January 9-13, 2025, is available at https://zenodo.org/records/15537678.

Bias

In-person

Topic and Knowledge-Base Interdisciplinarity in Manuscripts Submitted to Physical Science Journals vs Editorial Decision and Reviewer Positivity

Sidney Xiang,[1] Daniel M. Romero,[1,2,3] Misha Teplitskiy[1]

Objective Prior literature on interdisciplinary research evaluation does not include prepublication manuscripts or account for differential evaluation over multiple dimensions of interdisciplinarity. Based on management science literature, we introduce 2 novel dimensions: topic interdisciplinarity, which may incur evaluation penalties by cutting across disciplinary standards and threatening symbolic boundaries, and knowledge-base interdisciplinarity, which may incur benefits by combining larger pools of nonredundant information. Evaluations may also depend on how these dimensions align with each other and the audience. Our study tested these hypotheses using a large-scale dataset of journal submissions.

Design We performed a case-control study using administrative data from the Institute of Physics Publishing, a major STEM publisher. The data included 128,950 manuscripts (accepted and rejected) submitted to 62 physical science journals from 2018 to 2022. We calculated topic

interdisciplinarity with concept tags of the manuscript's associated OpenAlex record[1] and knowledge-base interdisciplinarity with concept tags of the manuscript's references. For each type, we quantified interdisciplinarity using the integration index,[2] ranging from 0 to 1. Logistic regression was used to model the relationship between both types of interdisciplinarity and 2 evaluation outcomes: final decision and review positivity, controlling for journal, manuscript type, year, number of authors, number of authors' prior citations, number of authors' prior publications, and number of references. We retained observations with both interdisciplinarities and all covariates.

Results The marginal effect of a 1-SD increase in knowledge-base interdisciplinarity was a 0.9 (0.7) percentage point higher acceptance (positive review) probability, while a 1-SD increase in topic interdisciplinarity corresponded to a 1.2 (0.4) percentage point lower acceptance (positive review) probability. Interactions revealed that high knowledge-base interdisciplinarity attenuated the negative relationship between topic interdisciplinarity and acceptance. We did not see a symmetric effect when high knowledge-base interdisciplinarity was accompanied by high topic interdisciplinarity. When disaggregating model results by journal interdisciplinarity according to the publisher's 2022 product catalogue,[3] journals categorized by the publisher as interdisciplinary or multisubject had positive associations between both types of interdisciplinarity and evaluation outcomes, whereas monodisciplinary journals had negative associations. Our results were robust to the inclusion and transformation of manuscript and team covariates.

Conclusions Our findings indicate that different dimensions of interdisciplinarity have different relationships to evaluation and that, for successful publication, authors must attend to alignment between types of interdisciplinarity within the manuscript and between the manuscript and journal audience. Limitations include the scope of journals in our dataset (STEM only) and the reliance on matching submissions to OpenAlex records.

References

1. Priem J, Piwowar H, Orr R. OpenAlex: a fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*. Preprint posted online May 4, 2022. doi:10.48550/ARXIV.2205.01833

2. Porter AL, Cohen AS, David Roessner J, Perreault M. Measuring researcher interdisciplinarity. *Scientometrics*. 2007;72(1):117-147. doi:10.1007/s11192-007-1700-5

3. IOP Publishing. Product catalogue 2022. https://ioppublishing.org/wp-content/uploads/2022/03/IOPP-Catalogue-2022.pdf

[1]University of Michigan School of Information, Ann Arbor, MI, US, tepl@umich.edu; [2]University of Michigan Department of Electrical Engineering and Computer Science, Ann Arbor, MI, US; [3]University of Michigan Center for the Study of Complex Systems, Ann Arbor, MI, US.

Conflict of Interest Disclosures None reported.

The Influence of Promotional Language on Evaluations of Biomedical Literature: A Randomized Controlled Trial

Neil Millar,[1] Brian Budgell[2]

Objective Promotional language (hereafter, *hype*) in scientific writing has increased significantly over the past 40 years.[1] Analyses of grant proposals have found the incidence of hype to be positively associated with funding success[2]; however, evidence of causality is still lacking. Furthermore, an underpowered pilot study suggested that hype may bias clinicians' perceptions and evaluations of evidence.[3] This study tested the hypothesis that students' appraisal of biomedical research abstracts is influenced by hype.

Design A double-blind randomized controlled trial was conducted to assess readers' evaluation of hyped vs nonhyped abstracts. Four structured abstracts were selected from leading spinal care journals, controlling for length and absence of hype. Hyped abstracts were modified to include 6 common hype terms, such as "carefully designed" to highlight methodological rigor, "this is the first study to" to emphasize novelty, "convincing evidence" to amplify results, and "conducted by an experienced radiologist" to emphasize researcher competence. Inclusion criteria were that participants were second-year or higher chiropractic students recruited from institutions in the UK, US, and Canada. A blinded researcher randomly selected folders for students to evaluate. Participants rated abstracts on a 10-point scale, using 4 criteria: implementation likelihood, rigor, novelty, and researcher competence. A linear mixed-effects model was used to assess whether evaluations were influenced by hype, accounting for variability across participants and abstracts.

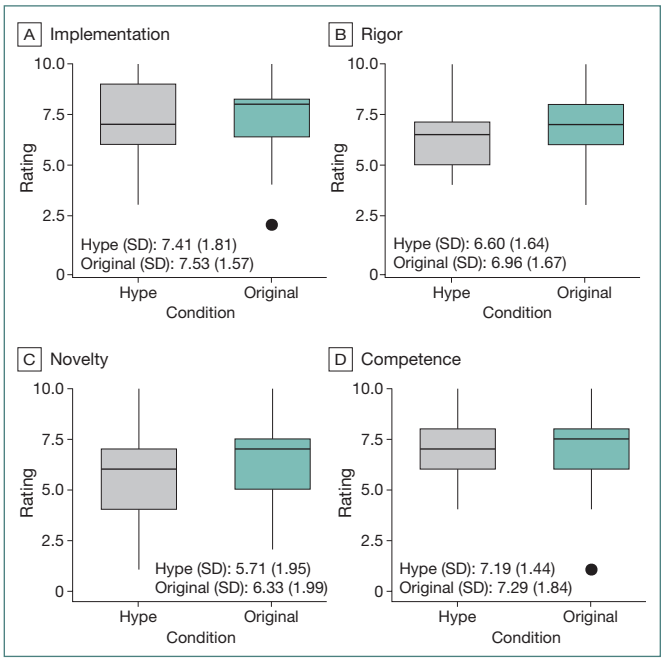
Results This ongoing study began on April 2, 2024. As of January 2025, 36 students at 3 institutions had completed the rating task. Mixed-effects modeling, including interactions between condition and criterion, showed no significant differences between hyped and original abstracts across all evaluation criteria (estimate, 0.067; $P > .05$) (**Figure 25-0927**). While novelty received consistently lower ratings (estimate, −1.479; $P < .001$), the influence of hype did not vary significantly by criterion, as interaction terms were nonsignificant (eg, condition × novelty interaction [estimate, 0.517; $P > .05$]), indicating minimal measurable effect of promotional language.

Conclusions Within this experimental paradigm, hype did not appear to bias readers' evaluation of the scientific merit of the literature. These findings are discussed in relation to evidence-based medicine and bias on the part of researchers, reviewers, editors, and other stakeholders. Limitations of the study include the small sample size reported to date and restriction of the experimental texts to abstracts.

References

1. Millar N, Batalo B, Budgell B. Trends in the use of promotional language (hype) in National Institutes of Health

Figure 25-0927. Ratings by Condition and Criterion[a]



[a]Annotated with mean rating and SD.

funding opportunity announcements, 1992-2020. *JAMA Netw Open*. 2022;5(11):e2243221. doi:10.1001/jamanetworkopen.2022.43221

2. Qiu HS, Peng H, Fosse HB, Woodruff TK, Uzzi B. Use of promotional language in grant applications and grant success. *JAMA Netw Open*. 2024;7(12):e2448696. doi:10.1001/jamanetworkopen.2024.48696

3. Millar N, Budgell B. Impact of hype on clinicians' evaluation of trials—a pilot study. *J Can Chiropr Assoc*. 2023;67(1):38-49.

[1]Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Ibaraki, Japan; [2]Canadian Memorial Chiropractic College, Toronto, Ontario, Canada, bbudgell@cmcc.ca.

Conflict of Interest Disclosures None reported.

Funding/Support This study was supported by grant 21K02919 from the Japan Society for the Promotion of Science.

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.

Unravelling the Spin and Selective Reporting in Medical AI Research

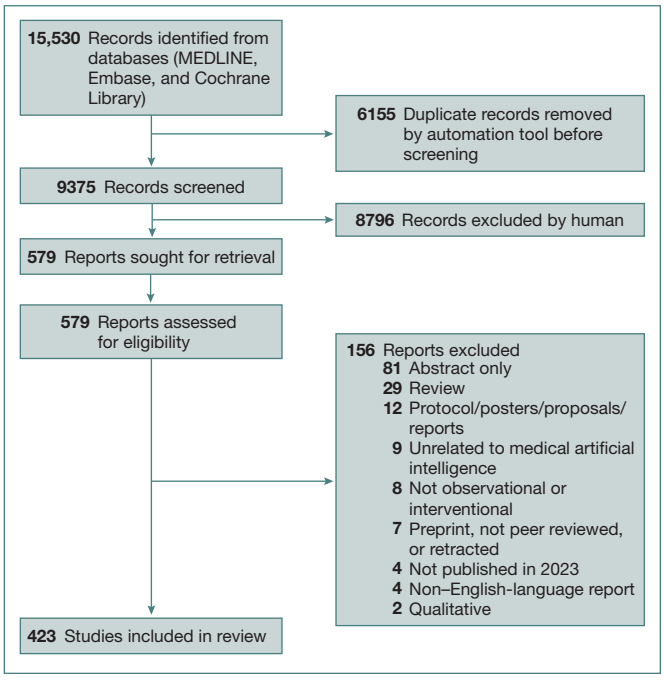
Vincent Yuan,[1] Aidan Christopher Tan[2]

Objective Our study aimed to determine the prevalence, types, and facilitators of "spin" and other selective reporting practices in recently published medical artificial intelligence (AI) research.

Design This was a cross-sectional meta-research study of interventional and observational studies related to medical AI. Medical AI was defined as computational models (eg, machine learning, deep learning, natural language processing) that amplify or augment health care–related tasks (eg, diagnosis, prognosis, risk stratification).[1] Studies were identified by searching MEDLINE, Embase, and the Cochrane Library. Studies were included if they were observational or interventional studies related to medical AI published in 2023, English-language, and full text. Spin was defined as the distortion of findings to present the performance of a medical AI tool as more beneficial than it appears. Examples of spin include misrepresentation or omission of significant or nonsignificant results and selective focus on abstract conclusions. Other study characteristics investigated included the journal, academic qualifications of first authors, funding, and competing interests. The features of spin were derived from coding manuals of Boutron et al[2] and Andaur Navarro et al.[3] Spin was categorized as either misleading interpretation, referring to an overemphasis of results, or misleading transportability, referring to an exaggeration of a model's clinical applicability. Instances of spin were abstracted by manually reviewing the Results and Discussion sections of abstracts and body and extracting the data into a prepiloted data collection form. This form was modified from SPIN-Prediction Models to exclude or include certain indicators, with the addition of custom items based on previous spin studies. Data were collected by 1 reviewer, and a second reviewer independently recollected the data in instances of uncertainty.

Results A total of 423 studies were included (**Figure 25-0950**). Slightly more than one-third (35% [n = 149]) of studies described AI as superior to the comparator. Nine percent of studies (n = 38) contained leading words in the title, and more than half (56% [n = 236]) selectively focused abstract conclusions on significant results. In almost half

Figure 25-0950



(48% [n = 203]) of study abstracts, qualifiers (eg, *might*, *very*) were present to subtly suggest stronger findings than likely warranted. The majority of studies (71% [n = 300]) contained strong statements (eg, *clearly shows*) in the body to emphasise the AI model's performance, accuracy, or effectiveness. Twenty-four percent (n = 101) of studies made recommendations for clinical practice in the body without external validation in the same study.

Conclusions Our findings suggest that selective reporting practices, including spin, are prevalent in medical AI research. These practices can misinform readers of the technical performance and real-world applicability of AI tools. This is especially crucial for titles and abstracts because these sections often influence first impressions and inform clinical decision-making.

References

1. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. 2021;8(2):e188-e194. doi:10.7861/fhj.2021-0095

2. 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

3. Andaur Navarro CL, Damen JAA, Ghannad M, et al. SPIN-PM: a consensus framework to evaluate the presence of spin in studies on prediction models. *J Clin Epidemiol*. 2024;170:111364. doi:10.1016/j.jclinepi.2024.111364

[1]School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia, vincent.yuan1@student.unsw.edu.au; [2]Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia.

Conflict of Interest Disclosures None reported.

A Comparison of Self-Acknowledged Limitations With Risk-of-Bias Assessments From Systematic Reviews

Joe D. Menke,[1] Mengfei Lan,[1] Halil Kilicoglu[1]

Objective The importance of self-acknowledged limitations (SALs) has long been recognized.[1] Prior work detecting[2] and categorizing[3] limitation statements using natural language processing (NLP) has made large-scale assessments of limitation reporting possible. By comparing SALs in research articles with limitations identified by systematic reviewers, this study aimed to evaluate the quality and coverage of SALs in relation to risk of bias.

Design We constructed a dataset pairing SALs with limitations discussed by reviewers through risk-of-bias assessments. Articles from the Cochrane Database of Systematic Reviews available in the open-access subset of PubMed Central were downloaded in September 2024. For papers cited in these reviews, full texts were downloaded when available, and existing NLP models[3] were used to extract and classify sentences discussing limitations (December 2024). Identifying information and risk-of-bias assessments for cited articles were extracted from reviews' Characteristics of Included Studies tables when a Risk of Bias section was present; reviews without this section were excluded. For living reviews, duplicate assessments were removed to retain only 1 assessment per cited article across versions of a living review. Risk-of-bias statements, originally in tabular form (eg, "Allocation concealment [Selection bias] High risk CBA; no allocation concealment done."), were rephrased into natural language using a large language model (Llama3.3-70B). Low-risk components were excluded, retaining only high- and unclear-risk items. A random subset was manually reviewed to check for systematic issues, which did not reveal concerning patterns or hallucinations. Paraphrased information from reviews and SALs from cited papers were classified using NLP models that recognized limitations sentences and types.[3] We analyzed and compared the distribution of categorized limitations mentioned in both groups in an effort to assess the quality of SALs.

Results From 145 reviews, we extracted and classified limitations statements for 803 cited articles (5.54 articles per review). From risk-of-bias tables, 3172 limitations were extracted (3.95 per cited article). From 750 cited articles, 136,290 sentences were processed and classified (170 per article), yielding 5616 SAL sentences (7.49 per article). No SALs were found in 53 articles. Categorized data are reported in **Table 25-1017**. Outcome measures were the most common SAL (69.12% of articles), while reviewers noted missing data as the most common high-risk limitation (33.75% of articles) and blinding as the most common unclear risk (48.19% of articles). On average, 56.67% of reviewer-identified limitations were not self-acknowledged, including 38.10% of high-risk and 53.01% of unclear-risk limitations.

Conclusions We present categorized limitations data and compared SALs with reviewer-derived limitations, showing that more than one-third of high-risk limitations are not self-acknowledged. Our findings may be used to better understand the quality of existing self-acknowledged limitations as well as highlight categories that are potentially underreported by authors.

References

1. Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol*. 2007;60(4):324-329. doi:10.1016/j.jclinepi.2006.09.011

2. Kilicoglu H, Rosemblat G, Malički M, Ter Riet G. Automatic recognition of self-acknowledged limitations in clinical research literature. *J Am Med Inform Assoc*. 2018;25(7):855-861. doi:10.1093/jamia/ocy038

3. Lan M, Cheng M, Hoang L, Ter Riet G, Kilicoglu H. Automatic categorization of self-acknowledged limitations in randomized controlled trial publications. *J Biomed Inform*. 2024;152:104628. doi:10.1016/j.jbi.2024.104628

[1]School of Information Sciences, University of Illinois Urbana-Champaign, Champaign, IL, US, jmenke2@illinois.edu.

Conflict of Interest Disclosures None reported.

Table 25-1017. Comparison of Limitations by Source and Risk Category

Limitation category	Percentage						
	Self-acknowledged	Mentioned by reviewers (any risk)	High risk	Unclear risk	Any risk not self-acknowledged[a]	High risk not self-acknowledged[a]	Unclear risk not self-acknowledged[a]
Study design	18.06	8.72	3.36	5.35	78.57	77.78	79.07
Population	61.64	39.10	24.03	22.04	32.48	24.87	37.29
Setting	7.60	0.25	0.25	0	100	100	NA
Intervention	43.34	0.87	0.25	0.62	0	0	0
Control	11.33	3.99	1.62	2.37	84.38	84.62	84.21
Outcome measures	69.12	44.33	23.41	34.37	26.69	27.13	26.09
Missing data	36.11	56.91	33.75	29.27	60.39	53.87	66.81
Underpowered study	44.08	4.11	2.49	1.62	42.42	35.00	53.85
Randomization	11.58	24.03	8.47	17.06	82.38	73.53	86.86
Blinding	16.56	61.39	28.64	48.19	78.50	70.87	82.17
Study duration	16.44	0.87	0.75	0.12	28.57	16.67	100
Statistical analysis	18.18	5.60	2.37	3.24	77.78	89.47	69.23
Funding	3.11	1.99	0.87	1.12	93.75	100	88.89
Generalization	38.36	5.23	1.99	3.36	66.67	93.75	51.85

Abbreviation: NA, not applicable.

The table reports the percentage of research articles in which each limitation was self-acknowledged by authors, referenced by systematic reviewers (at any risk level [high or unclear]), and classified specifically as high risk or unclear risk. Also reported is the percentage of articles in which reviewers identified a limitation (any, high, or unclear), but it was not self-acknowledged by the original authors.

[a]Lower values indicate risks were acknowledged more frequently.

Funding/Support This work was supported by the National Library of Medicine of the National Institutes of Health (NIH; R01LM014079).

Role of the Funder/Sponsor The funder had no role in the study design or in the collection, analysis, or interpretation of data; writing of the report; or decision to submit the abstract for publication.

Additional Information The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Assessment of Spin Among Diagnostic Accuracy Meta-Analyses Published in Top Pathology Journals: A Systematic Review

Griffin Hughes,[1] Andrew Tran,[1] Sydney Marouk,[1] Eli Paul,[1] Matt Vassar[1,2]

Objective The overinterpretation of research findings—otherwise known as "spin"—is commonly assessed in clinical trials and synthesis designs of interventional evidence.[1,2] However, spin in diagnostic evidence has received less attention and has fewer published works.[3] Therefore, we conducted a systematic review aimed to assess spin in pathology's diagnostic accuracy meta-analyses.

Design Using systematic review methodology, we searched PubMed (MEDLINE) on September 21, 2023, for diagnostic test accuracy meta-analyses published in top 20 pathology journals. To be included for data extraction, studies must have (1) been a systematic review with meta-analysis of primary studies identified through a literature search, (2) pooled diagnostic accuracy effects for a quantitative summary effect estimation, and (3) been published in a top 20 pathology journal, defined by Google Scholar Metrics' h5-index. We excluded studies that (1) did not represent systematic reviews with meta-analyses; (2) did not quantitatively pool diagnostic accuracy measures; (3) were primarily prognostic in their aims; (4) were published before the publication of the Quality Assessment of Diagnostic Accuracy Studies quality tool; or (5) represented publication types such as erratum, retractions, or corrigendum. We derived 10 items of actual spin and 9 items of potential spin from previously published methodology assessing overinterpretation of results in quantitative diagnostic synthesis.[3] Studies with conclusions interpreted as positive were assessed for actual spin, whereas all studies were assessed for potential spin. Authors screened and extracted relevant data from sample studies in a masked duplicate fashion to reduce extraction errors. A third author was available to resolve discrepancies. We conducted descriptive statistics, including frequencies and percentages, using R in RStudio.

Results We identified 207 articles from PubMed for potential inclusion. After screening, 55 abstracts and full texts were available for full data extraction, with 80% (44 of 55) having positive conclusions germane to accuracy or clinical implications. At least 1 item of spin was present in every included article. Most positive conclusions in abstracts (75% [33 of 44]) and full texts (79.6% [35 of 44]) did not adequately reflect pooled estimates (**Table 25-1166**). A total of 34.5% of studies (19 of 55) used nonrecommended statistical approaches for pooling accuracy measures,

Table 25-1166. Frequency of Actual Overinterpretation Items of Included Studies

Actual overinterpretation item and description[3]	Presence of actual overinterpretation in abstracts	No. (%) (N = 44)	Presence of actual overinterpretation in full texts	No. (%) (N = 44)
A.1: "Positive conclusion, not reflecting the reported summary accuracy estimates"[a]	Yes	33 (75.0)	Yes	35 (79.6)
	No	11 (25.0)	No	9 (20.5)
A.2: "Positive conclusion, not taking high risk of bias and/or applicability concerns into account"	Yes	10 (22.7)	Yes	14 (31.8)
	No	34 (77.3)	No	30 (68.2)
A.3: "Positive conclusion, not taking heterogeneity into account"	Yes	39 (88.6)	Yes	9 (20.5)
	No	5 (11.4)	No	35 (79.5)
A.4: "Positive conclusion, focusing on the results of primary studies favoring the diagnostic accuracy of the test instead of the meta-analysis results"	Yes	1 (2.3)	Yes	0
	No	43 (97.7)	No	44 (100.0)
A.5: "Positive conclusion, selectively focusing on a selection of subgroups, tests, or accuracy estimates, while others were evaluated as well"	Yes	4 (9.1)	Yes	0
	No	40 (90.9)	No	44 (100.0)
A.6: "Positive conclusion, inappropriately extrapolated to a wider population or setting"	Yes	0	Yes	0
	No	44 (100.0)	No	44 (100.0)
A.7: "Positive conclusion, inappropriately extrapolated as surrogates for improvement in patient important outcomes"[a]	Yes	1 (2.3)	Yes	1 (2.3)
	No	43 (97.7)	No	43 (97.7)
		No. (%) (N = 55)		NA
A.8: "Stronger conclusion in abstract"[b]	Yes	20 (36.4)	NA	NA
	No	35 (63.4)	NA	NA
		No. (%) (N = 11)		No. (%) (N = 11)
A.9: "Conclusion claiming test equivalence or superiority based on indirect comparisons"[c]	Yes	0	Yes	8 (72.7)
	No	11 (100.0)	No	3 (27.3)
A.10: "Conclusion claiming test equivalence or superiority without performing statistical comparisons or claiming test equivalence for nonstatistically significant results"[c]	Yes	7 (63.6)	Yes	3 (27.3)
	No	4 (36.4)	No	8 (72.7)

Abbreviation: NA, not applicable.

[a]A.1 to A.7 were only applicable to abstracts and full texts with an overall classification of positive or positive with qualifier.

[b]A.8 was applicable to abstracts regardless of overall classification.

[c]A.9 and A.10 were only applicable to abstracts and full texts that made comparative statements to other index tests.

reflecting potential spin. All 55 studies reported CIs within their full text for precision interpretation.

Conclusions Diagnostic test accuracy meta-analyses published within the top 20 pathology journals consistently overinterpret their findings. Authors should ensure that their findings are properly contextualized within predetermined diagnostic performance criteria. Additionally, authors should ensure that summary estimates are pooled using appropriate and robust methods that maintain the correlated nature of sensitivity and specificity.

References

1. 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

2. Yavchitz A, Ravaud P, Altman DG, et al. A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *J Clin Epidemiol*. 2016;75:56-65. doi:10.1016/j.jclinepi.2016.01.020

3. McGrath TA, Bowdridge JC, Prager R, et al. Overinterpretation of research findings: evaluation of "spin" in systematic reviews of diagnostic accuracy studies in high-impact factor journals. *Clin Chem*. 2020;66(7):915-924. doi:10.1093/clinchem/hvaa093

[1]Office of Medical Student Research, Oklahoma State University Center for Health Sciences, Tulsa, OK, US, griffinhughesresearch@gmail.com; [2]Department of Psychiatry and Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, OK, US.

Conflict of Interest Disclosures Matt Vassar receives 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 & Technology, and internal grants from Oklahoma State University Center for Health Sciences outside the present work. All other authors have nothing to disclose.

Study Hypotheses and Results From Superiority and Noninferiority Randomized Clinical Trials

Yuanxi Jia,[1] Yiwen Jiang,[2] Karen A. Robinson,[3] Jinling Tang[2]

Objective In embarking on randomized clinical trials (RCTs), researchers might hypothesize that a more intensive treatment will be better than a less intensive treatment (positive hypothesis) or that a more intensive treatment will be similar or noninferior to a less intensive treatment (negative hypothesis). Researchers might design noninferiority RCTs (NI-RCTs) to support negative hypotheses and superiority RCTs (S-RCTs) to support positive or negative hypotheses. Regardless of hypotheses, S-RCTs and NI-RCTs should produce consistent results when assessing similar participants, interventions, controls, and outcomes (PICO). Systematic discrepancies in effect estimates between S-RCTs and NI-RCTs assessing similar PICO might suggest the impact of bias. This study aimed to compare effect estimates between S-RCTs and NI-RCTs assessing similar PICO. We hypothesized that (1) S-RCTs with positive hypotheses produced larger effect estimates than NI-RCTs with negative hypotheses and (2) S-RCTs with negative hypotheses produced similar effect estimates to NI-RCTs with negative hypotheses.

Design This was a meta-research study of 101 meta-analyses comparing a more intensive treatment with a less intensive treatment (placebo, sham treatment, no treatment, or lower dose of the same treatment) that were identified from the Web of Science in July 2024. In total, 494 RCTs were analyzed, including 157 NI-RCTs and 337 S-RCTs (169 with positive hypotheses and 168 with negative hypotheses). In each meta-analysis, S-RCTs were selected as the exposure group and NI-RCTs as the control group. The bias for blinding was assessed using the Cochrane Risk of Bias tool. The ratios of effect estimates (risk ratio, mean difference [transformed to odds ratio], and hazard ratio) between S-RCTs and NI-RCTs were calculated and then combined across meta-analyses to form a single estimate as the main outcome. The ratio of effect size (RES) was estimated among all RCTs and prospectively registered RCTs, the initial hypotheses of which were verified.

Results S-RCTs with positive hypotheses had higher effect size estimates than NI-RCTs (RES, 1.47; 95% CI, 1.28-1.70). When restricted to prospectively registered RCTs, the RES was 1.47 (95% CI, 1.20-1.82). Among RCTs rated as low risk of bias for blinding, the RES was 1.03 (95% CI, 0.74-1.44), while among those rated as high or unclear risk of bias for blinding, the RES was 1.83 (95% CI, 1.45-2.32). S-RCTs with negative hypotheses produced effect estimates similar to those of NI-RCTs (RES, 0.94; 95% CI, 0.85-1.04). When restricted to prospectively registered RCTs, the RES was 1.00 (95% CI, 0.80-1.25). Among RCTs rated as low risk of bias for blinding, the RES was 0.93 (95% CI, 0.71-1.22), while among those rated as high or unclear risk of bias for blinding, the RES was 1.14 (95% CI, 0.90-1.44).

Conclusions Researchers' hypotheses may bias the results of RCTs. Blinding should be emphasized to reduce bias from researchers' hypotheses. Systematic reviews and clinical practice guidelines are suggested to routinely assess the impact of researchers' hypotheses on clinical evidence.

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, yx.jia@nus.edu.sg; [2]Shenzhen Institute of Advanced Technology, Shenzhen, China; [3]School of Medicine, Johns Hopkins University, Baltimore, MD, US.

Conflict of Interest Disclosures None reported.

Funding/Support This study was supported by the Shenzhen Science and Technology Program (grant No. KQTD20190929172835662) from the Shenzhen Municipal Government, Guangdong Province, China, and the Outstanding Youth Innovation Fund from the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (grant No. E2G019).

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.

Biomedical Studies Published With Negative Results Over the Past Decade

Hannah Grace Varkey,[1] Florian P. Thomas,[1,2] Elli Gourna Paleoudis[3,4]

Objective Current scientific practices and culture are largely focused on producing high numbers of publications and positive or newsworthy data. Recently, some journals have dedicated space to the publication of negative results; nevertheless, negative results often remain unpublished, leading to duplication of efforts, lack of transparency, and publication bias. To investigate the evolution of negative data publications over a decade, we reviewed published studies that did not reach statistical significance or reported inconclusive findings or trends.

Design We searched for negative results in 300 publications per year from 2014 to 2024 using Web of Science's Medline. To limit selection bias, we selected the first 300 articles of each year based on the Web of Science publication date in chronological order, without regard to the scientific or treatment area. Additional requirements included clinical trial or other experimental design, presence of an abstract, and English language; case reports, commentaries, editorials, and case series were excluded. Records from each year were reviewed independently by 2 researchers to ensure reliability and accuracy. Any disagreements were resolved by a tie-breaking third researcher.

Results Over 10 years, we found 2 to 6 articles annually (**Figure 25-0879**) with solely negative results, with a downward trend (from 6 in 2014 to 3 in 2024). This assessment was applied to the primary outcome measure if and as identified. For the purposes of this project, negative results were defined as results that were inconclusive (ie, they did not prove a hypothesis or yield unexplained outcomes) or statistically nonsignificant based on the predefined measure

of statistical significance (usually a *P* value). Identified articles with negative outcomes were most commonly from the fields of cellular biology, biomechanics, and genetics.

Conclusions Recently, there has been increased attention to and support for publishing negative findings in traditional and new scientific journals and search engines (eg, BioNØT[1]) focused specifically on these results. The notion that all research, regardless of outcome, can contribute to scientific advancement is gaining traction among journal editors, funders, and the scientific community overall.[2] This is particularly true in the artificial intelligence and machine learning era, where models are created based on existing (ie, published) data and research is driven by large datasets.[3] Access to negative results would not only limit duplication of effort but also improve planning and interpretation of data. Thus, it may be surprising that publications with negative data remain few in number, and such results are often rejected by journals, forgotten as unpublished data, or buried in supplemental sections.

References

1. Agarwal S, Yu H, Kohane I. BioNØT: a searchable database of biomedical negated sentences. *BMC Bioinformatics*. 2011;12(1):420. doi:10.1186/1471-2105-12-420

2. Sansom O, Bogani D, Reichenbach L, Wells S. Negative equity—the value of reporting negative results. *Dis Model Mech*. 2024;17(8):dmm050937. doi:10.1242/dmm.050937

3. Taniike T, Takahashi K. The value of negative results in data-driven catalysis research. *Nat Catal*. 2023;6:108-111. doi:10.1038/s41929-023-00920-9

[1]Department of Neurology, Hackensack Meridian School of Medicine, Nutley, NJ, US; [2]Hackensack University Medical Center, Nutley, NJ, US; [3]Investigator Initiated Research Program and Support Services, Office of Research Administration, Hackensack Meridian Health Research Institute, Nutley, NJ, US; [4]Department of Medical Sciences, Hackensack Meridian School of Medicine, Hackensack Meridian Health, Nutley, NJ, US; elli.gournapaleoudis@hmhn.org.

Conflict of Interest Disclosures None reported.

Bibliometrics and Publication Metrics

In-person

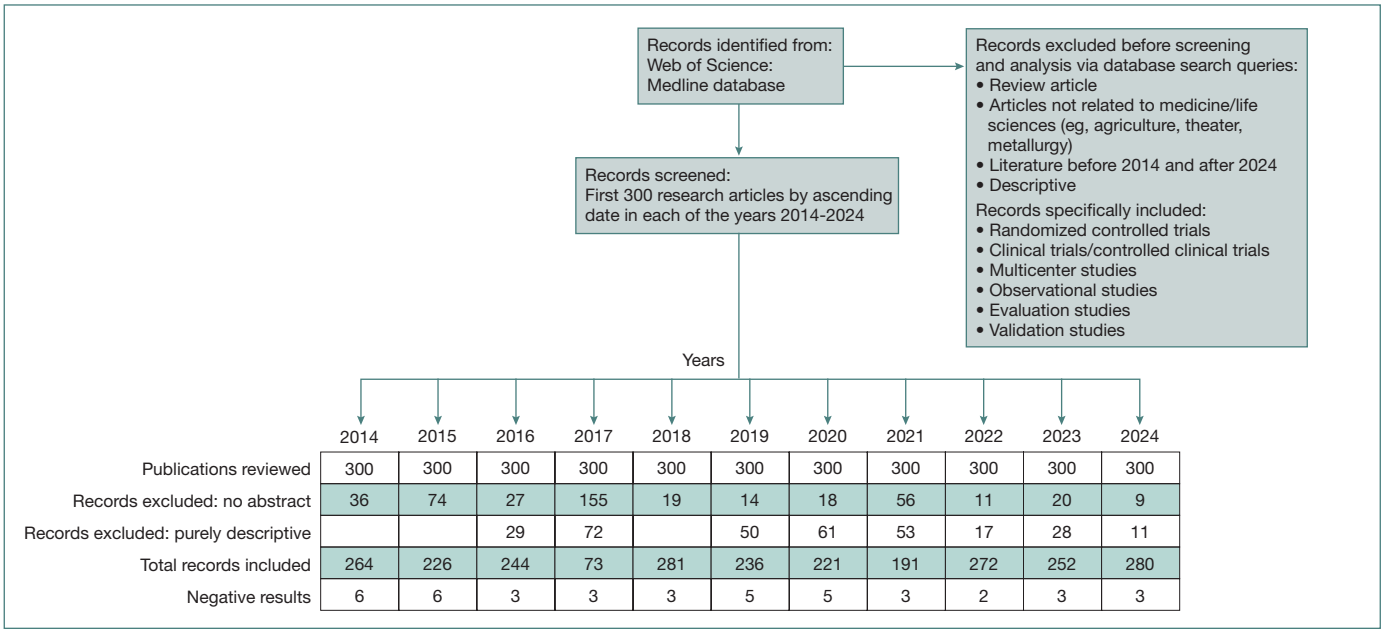
Review and Publication Times Across Journals Publishing on Health Policy

Kathryn A. Phillips,[1,2] Danea M. Horn[1,2]

Objective The move toward open science has focused attention on journal review and publication times,[1] particularly for health care research, where rapid dissemination is important for decision-making. We examined how academic journals that publish on health policy report and measure their review and publication times (RPTs) and assessed patterns across journals. Compared with earlier research that focused on article-level metrics for specialty journals,[2] we focused on journal-level metrics across a broad range of journals, enhancing generalizability.

Design We conducted a cross-sectional analysis of 57 academic journals that publish on health policy topics following STROBE guidelines. Institutional review board review was not required. We included journals with a Google h5-index greater than 23 and a Clarivate impact factor of 2 or higher as of November 18, 2024. Final inclusion was based on review by 3 journal editors. We examined 6 metrics (**Table 24-0820**) and used descriptive and statistical analyses to assess variation in the definition and reporting of RPTs. We compared RPTs across journal characteristics using the Wilcoxon rank sum test (with significance established at $P < .05$).

Figure 25-0879. Web of Science Search for Studies With Negative Results, 2014 to 2024



	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Publications reviewed	300	300	300	300	300	300	300	300	300	300	300
Records excluded: no abstract	36	74	27	155	19	14	18	56	11	20	9
Records excluded: purely descriptive			29	72		50	61	53	17	28	11
Total records included	264	226	244	73	281	236	221	191	272	252	280
Negative results	6	6	3	3	3	5	5	3	2	3	3

Table 24-0820. Descriptive Statistics for the Sample of Journals That Report at Least 1 Publication Timing Metric

Publication timing metric	No. of journals (N = 48)	No. of days	
		Mean (SD)	Median (IQR)
Time to first decision	38	13.9 (12.8)	10.0 (2.0-67.0)
Submission to decision, peer reviewed	28	80.9 (55.7)	60.5 (21.0-263.0)
Submission to final decision, peer reviewed	25	197.7 (67.7)	198.0 (38.0-314.0)
Submission to publication	7	164.7 (88.2)	150.0 (73.0-333.0)
Acceptance to publication, online	26	31.6 (39.2)	25.5 (2.0-205.0)
Acceptance to publication, print	5	124.6 (90.9)	105.0 (39.0-252.0)

Results Of the 57 journals analyzed, 9 (15.8%) did not publicly report any RPTs. Conversely, only 21 journals (36.8%) reported 3 or more metrics. There was wide variation in each reported RPT, with the full peer-reviewed publication process taking 5.6 to 11.9 months. Open access journals had faster median submission to final decision (peer-reviewed) times compared with hybrid and subscription journals (184.5 vs 228.0 days; $P = .03$). Highly selective journals (journals with high impact factors and low acceptance rates) had a faster median time to first decision than less selective journals (3.0 vs 13.0 days; $P < .001$). Within each RPT metric, there was considerable variation in reporting frequency and definitions. Journals often did not clarify whether the time to first decision included peer review and presented ranges or used qualifiers such as *typically* or *approximately* for post–peer review decision times. Few journals reported metrics by article type or provided metric time frames. Lastly, *mean* and *median* were used inconsistently.

Conclusions There was substantial variation across journals in whether and how they report RPTs and selective reporting and considerable variation in how timing metrics were defined. Publication speed is influenced by multiple factors, including review processes and publication frequency. The lack of standardized reporting practices limits transparency in academic publishing and suggests a need for journals to more consistently report RPTs and for RPTs to be included in standardized databases and tracking systems. This would better inform author submission decisions and inform future research and testing of RPT interventions using data from editorial systems. Improving reporting and transparency is especially important for research on urgent health policy topics and for early-career academics.[3]

References

1. Phillips KA. Open access publication at a crossroads—implications for researchers and beyond. *JAMA Health Forum*. 2024;5(10):e242914. doi:10.1001/jamahealthforum.2024.2914

2. Petrou C. Guest post—publishing fast and slow: a review of publishing speed in the last decade. The Scholarly Kitchen. November 8, 2022. Accessed November 8, 2024. https://scholarlykitchen.sspnet.org/2022/11/08/guest-post-publishing-fast-and-slow-a-review-of-publishing-speed-in-the-last-decade/

3. Andersen MZ, Fonnes S, Rosenberg J. Time from submission to publication varied widely for biomedical journals: a systematic review. *Curr Med Res Opin*. 2021;37(6):985-993. doi:10.1080/03007995.2021.1905622

[1]Center for Translational and Policy Research on Precision Medicine, Department of Clinical Pharmacy, University of California, San Francisco, CA, US, kathryn.phillips@ucsf.edu; [2]Institute for Health Policy Studies, University of California, San Francisco, CA, US.

Conflict of Interest Disclosures Kathryn A. Phillips is the founding editor in chief of *Health Affairs Scholar: Emerging and Global Health Policy*. No other conflicts were reported.

Funding/Support This abstract was partially supported by Health Affairs and Project HOPE.

Role of the Funder/Sponsor The funder had no role in the preparation, review, or approval of the abstract.

Funding Sources and the Online and Academic Impact of Cardiovascular Trials Published in Highest-Impact Journals

Farbod Zahedi Tajrishi,[1] Sina Rashedi,[2] Ashkan Hashemi,[3] Isaac Dreyfus,[4] Nicholas Varunok,[5] John Burton,[6] Seng Chan You,[7] Björn Redfors,[3,8] Gregory Piazza,[2] Joshua D. Wallach,[9] Lesley Curtis,[10] Sanjay Kaul,[11] David J. Cohen,[8,12] Roxana Mehran,[13] Flavia Geraldes,[14] Joseph S. Ross,[15] Jane Leopold,[2] Harlan M. Krumholz,[15] Gregg W. Stone,[13] Behnood Bikdeli[2,15]

Objective Altmetric scores capture online engagement with research, including social media mentions and news coverage, whereas citation counts represent academic impact. It remains uncertain whether the source of funding for randomized clinical trials (RCTs) correlates with the online and academic impact of RCTs overall and within the subsets that meet or do not meet their prespecified primary outcomes. We hypothesized that privately funded trials, particularly those with positive results, would be cited more often and have higher Altmetric scores than publicly funded trials.

Design We searched PubMed for cardiovascular RCTs published in *JAMA*, *The Lancet*, and *New England Journal of Medicine* between 2014 and 2019. This timeframe was selected to ensure sufficient follow-up for citation accrual and stability of Altmetric activity while capturing contemporary cardiovascular research. Trials were classified by funding source: public (ie, governmental), private (including for-profit industry and nonprofit organizations), or hybrid (with both public and private funding). RCT results were considered positive if they met their primary outcome or at least 1 of their coprimary outcomes and negative if they did not. Altmetric scores were obtained from Altmetric.com, and citation counts from Google Scholar (last accessed February 8, 2025). Both metrics were adjusted for publication year.

Results A total of 435 cardiovascular RCTs were analyzed, of which 130 (29.9%) were publicly funded and 197 (45.3%) were privately funded (**Figure 25-1147**). A total of 175 negative trials (40%) were identified. There was no significant difference across trials for Altmetric score based on the funding source (median [IQR] scores: public, 218.1 [112.3-465.6]; private, 215.1 [112.9-506.6]; and hybrid, 244.5 [127.3-555.8]; *P* = .49). Publicly funded trials were cited less frequently than privately funded or hybrid-funded trials (median [IQR] citations: public, 410.0 [226.5-724.0]; private, 599.0 [287.0-1022.0]; and hybrid, 607.0 [301.8-936.0]; *P* = .003). When exploring the findings based on the primary results of the RCTs, median (IQR) citations for positive vs negative publicly funded trials (402.5 [216.3-827.0] vs 419.5 [240.3-644.0] citations) and hybrid trials (614.5 [311.8-1173.3] vs 597.0 [300.5-808.3] citations) were substantively similar, whereas median (IQR) citations for privately funded trials (705.0 [298.5-1395.5] vs 383.0 [262.3-694.8] citations) were higher if they met the primary outcome (*P* for interaction < .001).

Conclusions Funding source was associated with subsequent citations, but not the Altmetric score. Publicly funded trials received significantly fewer citations than privately funded trials. Publicly and hybrid-funded trials showed similar citation patterns regardless of outcome, whereas privately funded trials were cited more when the primary outcome was met. These findings highlight the role of funding source in shaping the impact of cardiovascular RCTs.

[1]Tulane University School of Medicine, New Orleans, LA, US; [2]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US, bbikdeli@bwh.harvard.edu; [3]Weill Cornell Medicine, New York, NY, US; [4]David Geffen School of Medicine, University of California, Los Angeles (UCLA), Los Angeles, CA, US; [5]Vanderbilt University Medical Center, Nashville, TN, US; [6]Keck School of Medicine, University of Southern California (USC), Los Angeles, CA, US; [7]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea; [8]Cardiovascular Research Foundation, New York, NY, US; [9]Rollins School of Public Health, Emory University, Atlanta, GA, US; [10]Duke University and Duke Clinical Research Institute (DCRI), Durham, NC, US; [11]Cedars-Sinai Medical Center, Los Angeles, CA, US; [12]St Francis Hospital & Heart Center, Roslyn, NY, US; [13]Icahn School of Medicine at Mount Sinai, New York, NY, US; [14]The Lancet Group, London, UK; [15]Center for Outcomes Research and Evaluation (CORE), Yale School of Medicine, New Haven, CT, US.

Conflict of Interest Disclosures None reported.

Virtual

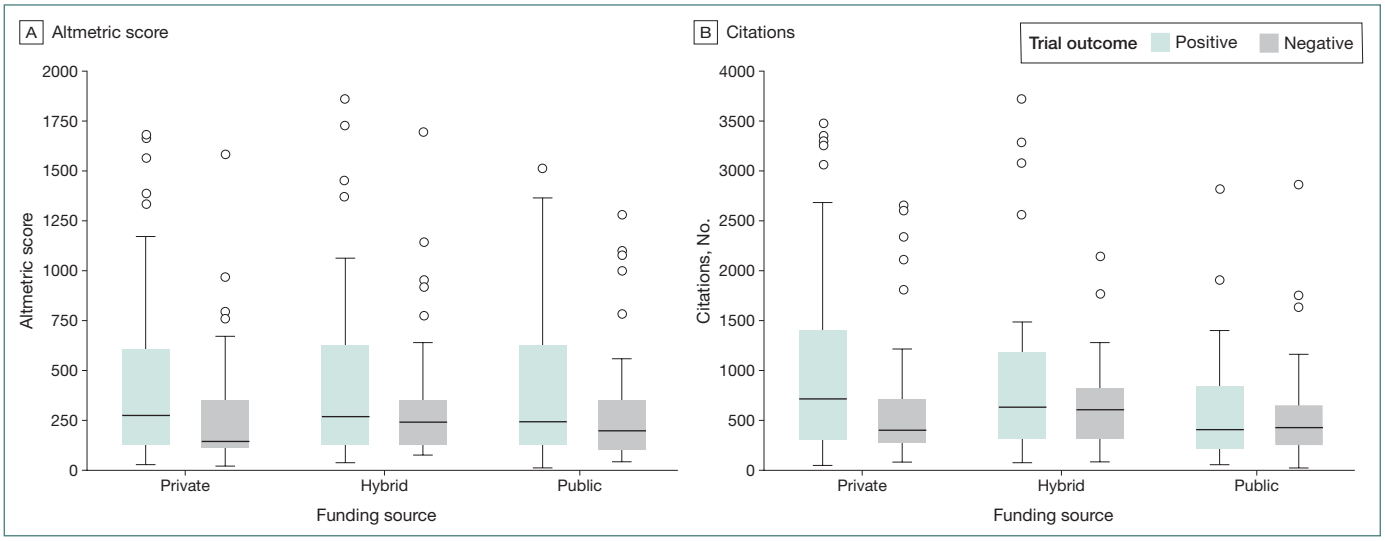
Publication Trends in Priority Epidemics According to the Sustainable Development Goals in Pharmaceutical Journals, 2000-2024

Julia Soto Rizzato,[1] Marcus Tolentino Silva,[2] David Moher,[3] Tais Freire Galvão[1]

Objective To assess publication trends in epidemics estimated to end by 2030 according to the Sustainable Development Goals (SDGs) in pharmaceutical journals in the past 25 years and the association of the adoption of these SDGs in 2015 with this trend. Assessments of the transition from Millennium Development Goals to SDGs are available,[1] but investigations are lacking on the impact of the adoption of these SDGs on publications of a specific target.

Design This cross-sectional study assessed all journals listed in the Pharmacology and Pharmacy category of the Journal Citation Reports. The primary outcome was the proportion of articles published addressing AIDS, tuberculosis, and malaria—epidemics targeted to end according to SDG 3.3. We consulted the Clarivate Web of Science Core Collection on December 19, 2024, to identify the total number of articles and the number of publications in each journal from 2000 to 2024 related to these epidemics. Data on country were collected from Web of Science metadata, which are based on authors' affiliations and then classified into low-, middle-, or high-income countries based on World Bank country classifications by income level. A paper was considered from a

Figure 25-1147. Funding Source, Engagement, and Citations, Stratified by Trial Outcome



The boxplot shows the distribution of citations in each subgroup, with the bottom of the box showing quartile 1; top of the box, quartile 3; and the horizontal line, quartile 2 (median). The whiskers show the bottom and top numbers still within 1.5 times the IQR, and circles indicate outlier values falling outside the whiskers.

low- or middle-income country if at least 1 author was from 1 of the countries in the metadata. An interrupted time series analysis was conducted to calculate the regression coefficient (β) and 95% CI of the proportion of papers (per 1000 publications) on SDG 3.3 before and after its adoption in 2015. Stata, version 14.2 was used for statistical analyses.

Results Three hundred fifty-five journals were included, which published 1,313,671 articles, of which 49,687 (3.8%) addressed SDG 3.3 from 2000 to 2024 and 37,521 (2.9%) were published by authors from high-income countries and 12 166 (0.9%) were published by authors from low- and middle-income countries. Between 2000 and 2015, there was an overall growing trend in published papers related to SDG 3.3 (β, 0.47; 95% CI, 0.18-0.76; P = .003), which was more pronounced in developed countries (β, 0.69; 95% CI, 0.38-1.00; P < .001), while the trend from developing countries was not significant (β, 0.21; 95% CI, −0.19 to 0.06; P = .30) (**Figure 25-0901**). After the adoption of SDGs in 2015, SDG 3.3 publications started to decrease until 2024 (β, −1.00; 95% CI, −1.54 to −0.47; P = .001) in a similar pattern in high-income countries (β, −1.40; 95% CI, −2.00 to −0.81; P < .001), whereas a nonsignificant increase was observed in lower-income countries (β, 0.05; 95% CI, −0.05 to −0.73; P = .89) in the same period.

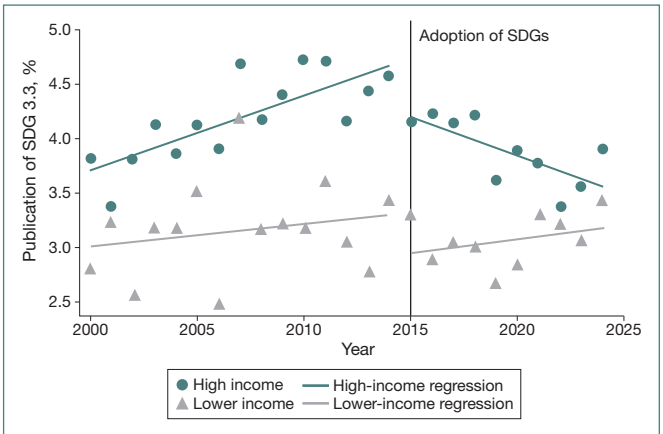
Conclusions The adoption of SDGs in 2015 did not seem to affect the publication priorities of pharmaceutical journals until 2024, which indicates that these measures may not have stimulated research efforts and publications. Even when considering that the results indicating research agenda priorities would take a few years to be published, the trend decreased until the end of the series, almost a decade after the adoption of SDGs.

Reference

1. Díaz-López C, Martín-Blanco C, De la Torre Bayo JJ, Rubio-Rivera B, Zamorano M. Analyzing the scientific evolution of the sustainable development goals. *Appl Sci*. 2021;11(18):8286. doi:0.3390/app11188286

[1]Faculdade de Ciências Farmacêuticas, Universidade Estadual de Campinas, Campinas, Brazil, taisgalvao@gmail.com; [2]Faculdade de Ciências de Saúde, Universidade de Brasília, Brasília, Brazil; [3]Centre for Journalology, Methods Centre of the Ottawa Hospital Research Institute, Ottawa, Canada.

Figure 25-0901. Publication Trends in the Sustainable Development Goal (SDG) Target 3.3 According to the Income Levels of the Authors' Countries



Conflict of Interest Disclosures David Moher and Tais Freire Galvão report being 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. The other authors declare no conflict of interest.

Funding/Support This study was funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (Finance Code 001, granted to Julia Soto Rizzato). Tais Freire Galvão receives a productivity scholarship from the National Council for Scientific and Technological Development (grant 313431/2023-00).

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

Science Journal Abstracts Misregistered in the Crossref Database

Qinyue Liu,[1] Yagmur Öztürk,[1] Cyril Labbé[1]

Objective Abstracts of scientific publications are widely used in scientometrics and text mining because they can be extracted in large quantities from bibliometric databases, and they convey important information about the publications. However, we found misregistered abstracts in Crossref published by *Science* and the *Proceedings of the National Academy of Sciences* (*PNAS*). We aimed at estimating the proportion of misregistered abstracts in our dataset and providing the textual similarity between the correct abstracts and the misregistered texts. A previous study also found misregistered metadata in Crossref.[1]

Design Having recently worked on citation analysis, we developed a script to automatically collect a dataset of citations extracted from articles published between 2016 and 2024 in *The Lancet, Cell*, and *Joule*. Each citation context is paired with the cited abstract mainly extracted using Crossref API. This dataset includes 19,822 citation context-abstract pairs. We noticed that certain abstracts on Crossref are misregistered. For *Science*, a section that introduces and cites the current article, and for *PNAS*, a section called Significance, are misregistered as abstracts on Crossref. We developed a script to automatically find these cases from our dataset by searching for distinctive features (string "et al" for *Science* and "Significance" for *PNAS*). For the cases in *Science*, we also manually collected the correct abstracts, and used the Sentence-BERT (SBERT)[2] model to calculate the textual similarity between the misregistered abstracts and the correct abstracts.

Results Of the 821 abstracts of *Science* publications, 402 (48%) contained the distinctive features. We manually verified the results, and 400 of 402 abstracts were misregistered. The 2 correctly registered abstracts were for report types of publication where there was no other text to be misregistered. Of the 669 *PNAS* abstracts, 243 (36%) contained Significance at the beginning. Moreover, most of the misregistered abstracts in our dataset, as estimated, had

high textual similarity scores (cosine similarity) with the correct abstracts. The scores range from 0.58 to 0.99, with an average of 0.83. This phenomenon occurs because the misregistered abstracts summarize the article and mention key points, resembling the original abstracts.

Conclusions Metadata misregistration in large repositories like Crossref can compromise data quality. Although our method detected misregistered abstracts in a small portion of the dataset using distinctive features, its scalability requires validation with larger datasets.

References

1. Cioffi A, Coppini S, Massari A, et al. Identifying and correcting invalid citations due to DOI errors in Crossref data. *Scientometrics*. 2022;127:3593-3612. doi:10.1007/s11192-022-04367-w

2. Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-Networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics; 2019:3982-3992. doi:10.18653/v1/D19-1410

[1]Université Grenoble Alpes, French National Centre for Scientific Research, Grenoble INP, Laboratoire d'Informatique de Grenoble, Grenoble, France, qinyue.liu@univ-grenoble-alpes.fr.

Conflict of Interest Disclosures None reported.

Funding/Support The NanoBubbles project has received Synergy grant funding from the European Research Council, within the European Union's Horizon 2020 program, grant 951393.

Role of Funder/Sponsor The European Research Council 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.

Trends in Citation Impacts of Original Research in Major Cardiovascular Journals, 2008-2018

Younwoo Ki,[1,2] Chungsoo Kim,[3,4] Yuan Lu,[3] Joshua D. Wallach,[5] Behnood Bikdeli,[4,6,7] Milton Packer,[8,9] Harlan M. Krumholz,[4] Seng Chan You[1,2]

Objective Recent trends in cardiovascular journals suggest a growing emphasis on publishing guidelines, reviews, and statements.[1,2] Citations, often used to assess the scholarly impact of journals, also influence researcher evaluation and library purchasing decisions. To understand these changes and their implications, we examined the evolving contribution of original research to citation metrics in leading cardiovascular journals.

Design Our cross-sectional bibliometric study used Scopus to identify all articles published in the *Journal of the American College of Cardiology* (JACC), *Circulation* (CIRC), and *European Heart Journal* (EHJ) from 2008 to 2018, with associated citations. Using section headings on respective journal websites and article metadata (title, abstract, and pages), we manually classified articles into (1) original research articles (ORAs, including nonsystematic meta-analyses); (2) reviews as nonoriginal research articles (NORA-reviews, narrative or systematic); (3) guidelines, statements, and statistics as NORAs (NORA-GSS), and (4) uncitable items (brief reports, case reports, commentaries, and editorials). Uncitable items were excluded from analysis. Outcomes were (1) ORAs' and NORAs' mean 2-year citation counts, (2) proportion of 2-year citations by each category among total 2-year citations to all citable items, and (3) publication count and proportions of each category among total citable items.

Results Among 22,740 articles, 9117 (39.9%) were ORAs, 2286 (10.0%) were NORA-reviews, 909 (4.0%) were NORA-GSS, and 10,566 (46.2%) were uncitable items. From 2008 to 2018, Journal Impact Factors (JIFs) in each journal increased, but their 2-year mean (SD) citations of ORA plateaued (JACC, 24.9 [24.3] to 26.4 [28.9]; CIRC, 24.00 [26.3] to 23.4 [23.6]; and EHJ, 15.8 [16.9] to 25.2 [20.8]) (**Figure 25-0977**, A). The annual publication of ORAs among total citable items declined across all 3 journals, although to varying extents (counts with proportions: JACC, 333 [80.5] to 245 [65.8]; CIRC, 415 [79.2] to 267 [78.1]; and EHJ, 251 [87.9] to 165 [69.6]), reflecting the decreased proportion of ORA 2-year citations among total citations (proportion of 2-year citations: JACC, 71.9% [9249/12,858] to 55.1% [7561/13,728]; CIRC, 68.1% [10,657/15,646] to 55.8% [6708/12,027]; and EHJ, 70.1% [4023/5739] to 36.8% [5073/13,771]) (**Figure 25-0977**, B). Although the annual publications of NORA-GSS were small in portion (counts with proportions: JACC, 33 [7.5%] to 22 [5.7%]; CIRC, 40 [7.0%] to 47 [13.0%]; and EHJ, 9 [3.1%] to 29 [11.9%]), the proportion of their 2-year citations among the total substantially increased (proportion of 2-year citations: JACC, 10.6% [1362/12,858] to 20.0% [2749/13,728]; CIRC, 19.3% [3019/15,646] to 37.7% [4535/12,027]; and EHJ, 16.6% [955/5739] to 52.4% [7211/13,771]) (**Figure 25-0977**, B).

Conclusions Rising JIFs in major cardiology journals were associated primarily with an expanding share of nonoriginal content. Overreliance on JIF in academic evaluation may unconsciously steer journals toward prioritizing highly cited nonoriginal content, systematically undermining the value of primary research.

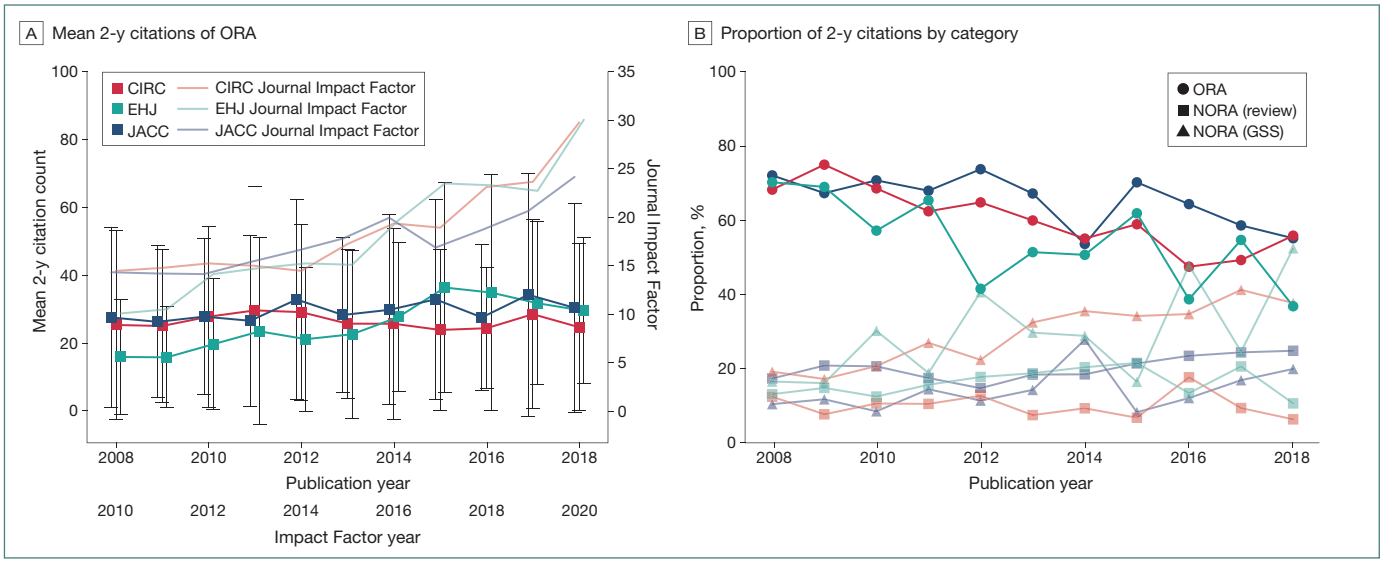
References

1. Genuis SJ. The proliferation of clinical practice guidelines: professional development or medicine-by-numbers? *J Am Board Fam Pract*. 2005;18(5):419-425. doi:10.3122/jabfm.18.5.419

2. Shuaib W, Khan MS, Shahid H, Valdes EA, Alweis R. Bibliometric analysis of the top 100 cited cardiovascular articles. *Am J Cardiol*. 2015;115(7):972-981. doi:10.1016/j.amjcard.2015.01.029

[1]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Korea; [2]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea, chandryou@yuhs.ac; [3]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven,

Figure 25-0977. Mean 2-Year Citation Count of Original Research Articles (ORAs) and Proportion of 2-Year Citation Counts by Article Categories in 3 Major Cardiovascular Journals From 2008-2018



A, Mean 2-year citation counts of ORAs. Solid lines show the annual mean number of citations within 2 years of publication for ORAs in *Circulation* (CIRC = red), the *European Heart Journal* (EHJ = green), and the *Journal of the American College of Cardiology* (JACC = blue); vertical bars represent ±1 SD. Transparent lines (right-hand y-axis) plot the corresponding Journal Impact Factors. B, Proportion of 2-year citations by article category: ORA (circle marker), review articles (NORA [review], square marker), and guidelines, statements, statistics (NORA [GSS], triangle marker). Values sum to 100% for every publication year. GSS indicates guidelines, statements, and statistics; NORA, nonoriginal research article.

CT, US; [4]Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, CT, US; [5]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, US; [6]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US; [7]Thrombosis Research Group, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US; [8]Baylor Heart and Vascular Institute, Baylor University Medical Center, Dallas, TX, US; [9]Imperial College London, London, UK.

Conflict of Interest Disclosures Several authors hold editorial positions at the *Journal of the American College of Cardiology* (*JACC*): Yuan Lu and Behnood Bikdeli as executive associate editors, Joshua D. Wallach and Seng Chan You as associate editors, Milton Packer as senior consulting editor, and Harlan M. Krumholz as editor in chief. Joshua D. Wallach reported receiving grants from Johnson & Johnson (through the Yale Open Data Access Project), Arnold Ventures, the National Institutes of Health, and the US Food and Drug Administration, as well as consulting fees from Hagen Berman Sobol Shapiro LLP and Dugan Law Firm APLC outside the submitted work. Seng Chan You reported being a chief technology officer of PHI Digital Healthcare and receiving grants from Daiichi Sankyo outside the submitted work.

Funding/Support This study was supported by a Severance Hospital Research Fund for Clinical Excellence (SHRC) (C-2024-0011).

Conflict of Interest

In-person

Psychiatry Editor in Chiefs' Publishing Practices in Their Own Journals

Justin N. Nguyen,[1] Robert T. Rubin[1,2]

Objective Several issues about scientific publishing, including integrity of peer review, honorary/ghost authorships, commercial influences, and inadequate disclosure of conflicts of interest, have been receiving increasing scrutiny.[1] The question arises as to how often journal editors in chief (EICs) publish research and review articles in their journals during their tenure compared with before and after their editorships.[2,3] If psychiatry EICs are publishing significantly more often during their tenure, might they be taking advantage of their position in their own journals? Our null hypothesis was that psychiatry EICs do not favor their own journals for self-publication.

Design As an exploratory study, EICs' publishing practices were examined for 7 general and specialty psychiatry journals (*Amer J Psychiatry*, *Acta Neuropsychiatrica*, *Biol Psychiatry*, *Can J Psychiatry*, *JAMA Psychiatry*, *J Clin Psychopharmacology*, and *Neuropsychopharmacology*) and 7 general medicine journals (*Am J Med*, *Ann Int Med*, *Br Med J*, *Can Med Assoc J*, *JAMA*, *Lancet*, and *N Engl J Med*). Data collection occurred between October 2021 and August 2024. For each journal, 1 to 6 psychiatry and 2 to 5 medicine EICs were identified, depending on lengths of tenure. From PubMed, publication lists were compiled for each EIC tenure period, 5 years prior, and 5 years following. Articles related to editors' responsibilities were omitted; only research and review articles (articles also appropriate for other journals) were included. Two-way analysis of variance with repeated measures was used to assess differences by specialty and time period (the repeated measure).

Results This study included 450 articles published in psychiatry journals and 96 published in general medicine journals. The mean and median articles per year for each time period are presented in **Table 25-0853**. Psychiatry editors published several times more research and review articles in their own journals than did medicine editors. In addition, psychiatry editors published 128% more frequently in their own journals while in office vs before compared with 24%

Table 25-0853. Rates of Self-Published Articles Per Year by Editors in Chief (EICs) Before, During, and After Their Tenure

Journal type	Mean (SD) [Median]		
	Before EIC tenure	During EIC tenure	After EIC tenure
Psychiatry	0.64 (0.78) [0.4]	1.46 (1.35) [1.44]	0.66 (0.89) [0.30]
Medicine	0.25 (0.40) [0]	0.31 (0.41) [0.21]	0.13 (0.25) [0]

more frequently for medicine editors. Analysis of variance indicated that the differences in both journal type (F = 10.33; df = 1,33) and time periods (F = 6.26; df = 2,66) were significant ($P < .003$ for both). Their interaction (F = 3.94; df = 2, 66) also was significant ($P < .02$). A total of 4 of the 25 psychiatry EICs accounted for most of this difference, with one EIC having published 9 times more frequently in their journal vs before and after their tenure.

Conclusions Psychiatry EICs published significantly more often when leading their journals, whereas medicine EICs did not. This finding has several important limitations. Selection of general medicine journals was subjective, and the medicine journals had higher Impact Factors, longer tenure of editors, and likely greater editorship responsibilities than the psychiatry journals. Articles with EICs as first, senior, and co-authors were included, and there may have been different author motivations for submission. We did not assess if the articles with EICs as authors included disclaimers about the role, or lack of role, of the EICs in editorial review and decision for these articles. Nevertheless, the increase in self-publishing occurred in a small number of psychiatry EICs. This should be further explored through study of a larger number of specialties, more closely matched journals, and more EICs.

References

1. Ioannidis JPA, Berkwits M, Flanagin A, Bloom T. Peer review and scientific publication at a crossroads. *BMJ*. 2023;382:1992. doi:10.1136/bmj.p1992

2. Helgesson G, Radun I, Radun J, Nilsonne G. Editors publishing in their own journals: a systematic review of prevalence and a discussion of normative aspects. *Learned Publishing*. 2022;35:229-240.

3. Liu F, Holme P, Chiesa M, AlShebli B, Rahwan T. Gender inequality and self-publication are common among academic editors. *Nat Hum Behav*. 2023;7:353-364. doi:10.1038/s41562-022-01498-1

[1]Community Memorial Hospital Psychiatry Residency, Ventura, CA, rtrubin@yahoo.com; [2]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine at UCLA, Los Angeles, CA.

Conflict of Interest Disclosures None reported.

Prevalence and Nature of Conflict of Interest Disclosures in Published Health Technology Assessment Reports

Miro Vukovic,[1] Ana Marušić[1]

Objective Conflict of interest (COI) disclosures are critical for the transparency of health technology assessment (HTA) reports, which inform health care policy. Previous studies have shown that COI disclosures in biomedical research and clinical guidelines are often incomplete or inconsistent, raising concerns about transparency and potential bias.[1-3] However, limited empirical evidence exists on how COIs are disclosed in HTA reports despite their critical role in health care decision-making. Our study aimed to address this gap by systematically examining COI disclosure practices across HTA publications.

Design We conducted a cross-sectional study of all English-language HTA reports (mini HTA, rapid reviews, and full HTA) from the International HTA Database (https://database.inahta.org/). Reports without accessible full text or not meeting HTA criteria were excluded. Data were extracted from reports published up to December 2023. We documented the presence of COI disclosure sections, their content, and use of standardized forms (eg, International Committee of Medical Journal Editors [ICMJE]) and reviewed adherence to the ICMJE COI reporting guidelines. Two independent reviewers conducted data extraction and coding, with disagreements resolved through discussion. Descriptive statistics were applied to report frequencies and proportions. Based on the study type, the STROBE guideline was followed for reporting.

Results Of 1218 included HTA reports, 598 (49.1%) were full HTAs, 174 (14.3%) were rapid reviews, and 78 (6.4%) were mini HTAs; 368 reports (30.2%) were excluded due to inaccessibility or duplication. Reports originated from 16 countries, with 330 (38.8%) from England. Overall, COI disclosure sections were present in 679 of 850 reports (79.9%). The lowest reporting was among mini HTAs (49 of 78 [62.8%]) compared with full HTAs (489 or 598 [81.8%]) and rapid reviews (141 of 174 [81.0%]). Only 98 reports (11.5% of included reports) used a standardized disclosure form, with the ICMJE form cited in 90. In another 8 reports, the EUnetHTA Declaration of Interest and Confidentiality Undertaking of Interest statement was used.

Conclusions While most HTA reports included COI disclosures, practices varied by report type and country. Standardized disclosure tools were rarely used. Further analysis is needed to support the development of harmonized and transparent COI policies across HTA-producing organizations.

References

1. Norris SL, Holmer HK, Ogden LA, Selph SS, Fu R. Conflict of interest disclosures for clinical practice guidelines in the National Guideline Clearinghouse. *PLoS One*. 2012;7(11):e47343. doi:10.1371/journal.pone.0047343

2. Norris SL, Burda BU, Holmer HK, Ogden LA. Conflicts of interest in clinical practice guideline development: a systematic review. *PLoS One*. 2011;6(10):e25153. doi:10.1371/journal.pone.0025153

3. Mandrioli D, Kearns CE, Bero LA. Relationship between research outcomes and risk of bias, study sponsorship, and author financial conflicts of interest in reviews of the effects of artificially sweetened beverages on weight outcomes: a systematic review of reviews. *PLoS One*. 2016;11(9):e0162198. doi:10.1371/journal.pone.0162198

[1]Centre for Evidence-Based Medicine, University of Split School of Medicine, Split, Croatia, miro.vukovic@mefst.hr.

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.

Development of a Tool for Addressing Conflicts of Interest in Trials (TACIT) for Use in Systematic Reviews

Andreas Lundh,[1,2,3] Isabelle Boutron,[4,5,6] Lesley A. Stewart,[7] Asbjørn Hróbjartsson[1,2]

Objective Industry funding and authors' financial conflicts of interest are common in clinical trials and associated with favorable trial conclusions and seemingly greater treatment effects. However, there has been little guidance on how to take funding and conflicts of interest into account in evidence synthesis. To address this, we developed the Tool for Addressing Conflicts of Interest in Trials (TACIT).

Design We established an international working group of 20 experts in trial and systematic review methods, conflicts of interest, and biomedical publishing. Four substudies informed tool development. First, a systematic review found that current study appraisal tools address funding and conflicts of interest superficially.[1] Second, an interview study of experienced trial researchers found that funding and conflicts of interest may unduly influence trials through various mechanisms but influence may be minimized if properly managed.[2] Third, a cross-sectional study of trials included in Cochrane reviews found that searching beyond trial publications using a structured information retrieval approach may improve identification of trial funding and researchers' conflicts of interest.[3] Fourth, an interview and questionnaire study of expected users piloting a preliminary version of TACIT found that users generally had a positive experience but suggested more guidance on ease of use and how to deal with insufficient information on conflicts of interest. The tool has been iteratively revised mainly based on feedback from the working group and expected users. A final consensus meeting was held in May 2025, where the result of substudies, themes for discussion, and a near-final version of the tool were presented. Based on the consensus meeting, the tool will be revised and made public together with a Microsoft Excel–based application after summer 2025.

Results Through a series of guidance questions, the tool facilitates a judgment of notable concern about conflicts of interest and the sufficiency of information that the judgment of concern was based on (**Table 25-0917**). The judgment involves 3 levels: trial funders, the trial's primary academic researchers, and overall trial level. The tool provides an overview of funding and conflicts of interest information for trials included in the systematic review, and it categorizes trials into trials with or without notable concern about conflicts of interest. Results of the assessments may be incorporated into review syntheses using subgroup or sensitivity analysis.

Conclusions TACIT provides a structured, coherent, and transparent assessment of conflicts of interest in trials included in systematic reviews. The tool may be used by systematic reviewers, guideline panels, and other users undertaking evidence syntheses.

Table 25-0917. Outline of TACIT[a]

Step	Domains and guidance questions[b]
Funders	
1.1	Record funding information
1.2	Are there important conflicts of interest for any of the trial funders?
1.3	Did any funders with important conflicts of interest have an important involvement in the trial?
1.4	How sufficient is the information that judgment of concern about conflicts of interest for the trial funders is based on?
Primary academic researchers	
2.1	Are there important conflicts of interest for any of the primary academic researchers?
2.2	How sufficient is the information that judgment of concern about conflicts of interest for the primary academic researchers is based on?
Overall trial level	
3.1	What is the concern about conflicts of interest for funders and for primary academic researchers?
3.2	How sufficient is the information that judgment of concern about conflicts of interest in the trial is based on?

Abbreviation: TACIT, Tool for Addressing Conflicts of Interest in Trials.

[a]TACIT aims to distinguish between trials with and without notable concern about conflicts of interest and the sufficiency of information that the judgment was based on. Final trial level judgment is (1) notable concern about conflicts of interest or no notable concern about conflicts of interest identified and (2) insufficient information or sufficient information.

[b]Important conflicts of interest indicates an interest in a particular magnitude or direction of the trial results; important involvement, influence on, or direct involvement in 1 or more stages of the trial (design, conduct, and analysis and reporting); insufficient information, not enough information for a reasonably confident judgment of concern about conflicts of interest (eg, missing information and/or indirect assessments necessary); sufficient information, enough information for a reasonably confident judgment of concern about conflicts of interest (ie, no relevant information missing and no indirect assessments); and primary academic researchers, a minimum of any academic researcher who is the first, last, or corresponding author or trial statistician of main trial publication.

References

1. Lundh A, Rasmussen K, Østengaard L, Boutron I, Stewart LA, Hróbjartsson A. Systematic review finds that appraisal tools for medical research studies address conflicts of interest superficially. *J Clin Epidemiol*. 2020;120:104-115. doi:10.1016/j.jclinepi.2019.12.005

2. Østengaard L, Lundh A, Tjørnhøj-Thomsen T, et al. Influence and management of conflicts of interest in randomised clinical trials: qualitative interview study. *BMJ*. 2020;371:m3764. doi:10.1136/bmj.m3764

3. Faltinsen E, Todorovac A, Boutron I, Stewart LA, Hróbjartsson A, Lundh A. A structured approach to information retrieval improved identification of funding and researchers' conflicts of interest in trials included in Cochrane

reviews. *J Clin Epidemiol*. 2023;161:104-115. doi:10.1016/j.jclinepi.2023.06.020

[1]Cochrane Denmark & Centre for Evidence-Based Medicine Odense (CEBMO), Department of Clinical Research, University of Southern Denmark, Odense, Denmark, alundh@health.sdu.dk; [2]Open Patient data Exploratory Network (OPEN), Odense University Hospital, Odense, Denmark; [3]Department of Respiratory Medicine and Infectious Diseases, Copenhagen University Hospital—Bispebjerg and Frederiksberg, Copenhagen, Denmark; [4]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAe, Centre for Research in Epidemiology and Statistics (CRESS), Paris, France; [5]Centre d'Épidémiologie Clinique, Hôpital Hôtel Dieu, Université Paris Cité, Paris, France; [6]Cochrane France, Paris, France; [7]Centre for Reviews and Dissemination, University of York, York, UK.

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.

Additional Information: The authors are members of the Tool for Addressing Conflicts of Interest in Trials (TACIT) Steering Group. We thank members of the TACIT working group and coauthors of TACIT substudies for their contribution to the development of TACIT.

Corporate Influence on Peer-Reviewed Research: Insights from BP's Deepwater Horizon Response

Marc-Andre Gagnon,[1] Blue Miaoran Dong[2]

Objective This study investigates BP's influence on peer-reviewed research in the wake of the 2010 Deepwater Horizon oil spill, with particular attention to how corporate oversight shaped the production and publication of scientific knowledge about the disaster and its aftermath. Despite the scale and impact of the incident, no academic work to date has systematically analyzed BP's internal documents—made public during litigation—or assessed their role in steering scholarly discourse through subtle or direct influence on academic journals.

Design Our study, grounded in 9 internal BP documents made public by The Downs Law Group (https://downslawgroup.com/bp-papers/), systematically examines BP's publication strategies. These documents reveal the extensive oversight of BP's legal and editorial teams in shaping the content of academic articles related to the spill. By analyzing publication plans, journal selection processes, thematic focuses, and the assertions made in these papers, we provide crucial insights into the challenges and manipulations inherent in peer review, scientific publication, and research dissemination.

Results The analysis highlights BP's direct role in the production of 37 peer-reviewed articles, as identified in one of the 9 leaked internal documents, "Scientific Literature Publication Tracker."[1] The body of literature under review was published between 2011 and 2024. Collaborating with institutions such as Florida International University, Oregon State University, SINTEF research company, and Commonwealth Scientific and Industrial Research Organisation, BP shaped research agendas, defined testing protocols, and influenced interpretations of results. A large portion of the articles (33) focused on "seeps," emphasizing natural seepage and exploration risks, while other themes included surf waters (18 articles), dispersants (12 articles), oil weathering (11 articles), and Corexit (6 articles). These studies often advanced conclusions that aligned with BP's strategic interests, such as promoting the ecological and economic benefits of chemical dispersants. Notably, less frequent topics like phototoxicity (3 articles) or biological impacts received comparatively limited attention. BP's editorial practices frequently framed findings within narrowly defined scopes to minimize ecological and legal accountability, steering the scientific narrative away from issues that could increase its liability. Among 37 articles, 17 did not disclose BP's involvement in the writing process, while 21 articles explicitly acknowledged BP's role. Although the 21 articles mentioned BP's funding and their receipt of a portion of the $500 million allocated through the Gulf of Mexico Research Initiative, they did not disclose any editorial review conducted by BP (**Table 25-0937**).

Conclusions The findings reveal vulnerabilities in the peer-review process, where corporate influence can compromise the objectivity and integrity of scientific research. BP's involvement demonstrates how editorial control and selective dissemination of findings can obscure accountability and shift public discourse. Enhanced transparency and robust safeguards are essential to protect the integrity of research and ensure it prioritizes public interest over corporate objectives.

Reference

1. BP. BP Scientific Literature Publication Tracker. Published online November 29, 2022. Accessed May 19, 2025. https://downslawgroup.com/bp-papers/

[1]The School of Public Policy, Carleton University, Ottawa, ON, Canada, ma.gagnon@carleton.ca; [2]School of Journalism and Communication, Carleton University, Ottawa, ON, Canada.

Conflict of Interest Disclosures None reported.

Funding/Support This research is financially supported by an Insight Grant awarded by the Social Sciences and Humanities Research Council of Canada (SSHRC) under the reference number 435-2021-0715, along with funding from the SSHRC Initiative for Digital Citizen Research under the reference number 1403-2021-0715.

Role of the Funder/Sponsor SSHRC did not play any role in the work described in the abstract.

Conflicts of Interest in Research Across Scholarly Disciplines

Helena Van Beersel Krejcikova,[1,2] Christoffer Bruun Korfitsen,[1,2] Lisa Bero,[3] Jason Dana,[4] David C. Dorman,[5] Quinn Grundy,[6] Ibo van de Poel,[7] Morten Rosenmeier,[8] Asbjørn Hróbjartsson,[1,2] Andreas Lundh[1,2,9]

Objective Conflicts of interest can undermine the trustworthiness of research if they are not adequately prevented or managed. Our objective was to (1) map and

Table 25-0937. List of 37 Articles Identified Through BP Scientific Literature Publication Tracker

Published paper	Publication year	Journal or publisher	Disclose Funding
Characterization and quantification of hydrocarbon seeps by means of subsea imaging	2015	IEEE	Yes
Assessing the impacts of oil-associated marine snow formation and sedimentation during and after the Deepwater Horizon oil spill	2016	*Anthropocene*	Yes
Recurrent oil sheens at the Deepwater Horizon disaster site fingerprinted with synthetic hydrocarbon drilling fluids	2013	*Environmental Science & Technology*	Yes
An adding-up test on contingent valuations of river and lake quality	2015	*Land Economics*	Yes
Natural seepage on the continental slope to the east of Mississippi Canyon in the northern Gulf of Mexico	2013	*Geochemistry, Geophysics, Geosystems*	Yes
Determination of sea-floor seepage locations in the Mississippi Canyon	2015	*Marine and Petroleum Geology*	No
An evaluation of select test variables potentially affecting acute oil toxicity	2016	*Archives of Environmental Contamination and Toxicology*	Yes
Acute aquatic toxicity studies of Gulf of Mexico water samples collected following the Deepwater Horizon incident	2015	*Chemosphere*	Yes
Factors affecting toxicity test endpoints in sensitive life stages of Native Gulf of Mexico species	2015	*Archives of Environmental Contamination and Toxicology*	Yes
A comparative assessment of the aquatic toxicity of Corexit 9500 to marine organisms	2018	*Archives of Environmental Contamination and Toxicology*	Yes
Characterization and environmental relevance of oil water preparations of fresh and weathered MC-252 Macondo oils used in toxicology testing	2017	*Science of The Total Environment*	Yes
Stability of dioctyl sulfosuccinate (DOSS) towards hydrolysis and photodegradation under simulated solar conditions	2014	*Science of The Total Environment*	Yes
Acute aquatic toxicity studies of Gulf of Mexico water samples collected following the Deepwater Horizon incident	2015	*Chemosphere*	Yes
Recommendations for advancing media preparation methods used to assess aquatic hazards of oils and spill response agents	2023	*Aquatic toxicology (Amsterdam, Netherlands)*	No
Use of TLM derived models to estimate toxicity of weathered MC252 oil based on conventional chemical data and the potential impact of unresolved polar components	2024	*Toxicology Mechanisms and Methods*	Yes
Depletion and biodegradation of hydrocarbons in dispersions and emulsions of the Macondo 252 oil generated in an oil-on-seawater mesocosm flume basin	2014	*Marine Pollution Bulletin*	Yes
Surface weathering and dispersibility of MC252 crude oil	2014	*Marine Pollution Bulletin*	Yes
Chemical comparison and acute toxicity of water accommodated fraction (WAF) of source and field collected Macondo oils from the Deepwater Horizon spill	2015	*Marine Pollution Bulletin*	Yes
Partitioning of PAHs between crude oil microdroplets, water, and copepod biomass in oil-in-seawater dispersions of different crude oils	2018	*Environmental Science & Technology*	Yes
Do oil droplets and chemical dispersants contribute to uptake of oil compounds and toxicity of crude oil dispersions in cold-water copepods?	2023	*Journal of Toxicology and Environmental Health, Part A*	No
Photo-enhanced toxicity of fluoranthene to Gulf of Mexico marine organisms at different larval ages and ultraviolet light intensities	2016	*Environmental Toxicology and Chemistry*	Yes
Photo-enhanced toxicity of two weathered Macondo crude oils to early life stages of the eastern oyster	2016	*Marine Pollution Bulletin*	Yes
Embryotoxicity of weathered crude oil from the Gulf of Mexico in mallard ducks	2011	*Environmental Toxicology and Chemistry*	No
Chronic effects of non-weathered and weathered crude oil and dispersant associated with the Deepwater Horizon incident on development of larvae of the eastern oyster	2016	*Environmental Toxicology and Chemistry*	Yes
A review of the emerging field of underwater mass spectrometry	2016	*Frontiers in Marine Science*	No
Fat tails and truncated bids in contingent valuation: an application to an endangered shorebird species	2016	*Ecological Economics*	No
Acute effects of non-weathered and weathered crude oil and dispersant associated with the Deepwater Horizon incident on the development of marine bivalve and echinoderm larvae	2016	*Deep Sea Research Part II: Topical Studies in Oceanography*	Yes
Chemical comparison and acute toxicity of water accommodated fraction (WAF) of source and field collected Macondo oils from the Deepwater Horizon spill	2015	*Marine Pollution Bulletin*	Yes
Effects of pollution on freshwater organisms	2012	*Water Environment Research*	No
Testing the sensitivity of stated environmental preferences to variations in choice architecture	2023	*Ecological Economics*	No
Productivity of waterbirds in potentially impacted areas of Louisiana in 2011 following the Deepwater Horizon oil spill	2018	*Environmental Monitoring and Assessment*	Yes
Near bottom acoustic and video measurements of the rise rate of methane bubbles in the Gulf of Mexico	2012	*The Journal of the Acoustical Society of America*	Yes
Transience and persistence of natural hydrocarbon seepage in Mississippi Canyon, Gulf of Mexico	2016	*Deep Sea Research Part II: Topical Studies in Oceanography*	Yes
Impacts of Macondo oil from Deepwater Horizon spill on the growth response of the common reed Phragmites australis: a mesocosm study	2014	*Marine Pollution Bulletin*	Yes
Underwater acoustic technology-based monitoring of oil spill: a review	2023	*Journal of Marine Science and Engineering*	No
Development and validation of a procedure for numerical vibration analysis of an oscillating wave surge converter	2016	*European Journal of Mechanics-B/Fluids*	No
Comparison of the acute toxicity of Corexit 9500 and household cleaning products	2015	*Human and Ecological Risk Assessment*	Yes

analyze conflict of interest policies of important international and national scholarly organizations within 5 selected scholarly disciplines and (2) describe cross-disciplinary differences and commonalities.

Design This was a cross-sectional study and content analysis of conflict of interest policies for research in economics, engineering, environmental toxicology and chemistry, law, and medicine. We used purposive sampling to include international and national policies in English from: (1) 100 top scholarly journals (20 per discipline) and the 20 largest publishers (both based on rankings in Clarivate's Journal Citation Reports 2022); (2) major journal and publisher associations and international multidisciplinary scholarly organizations (identified via targeted web search); (3) top universities (2 per continent based on ranking in Times Higher Education World University Rankings 2023); and (4) public funding agencies with the largest amount of annual funding (up to 2 per continent and discipline, identified via targeted web search). One researcher identified potentially eligible policies (April 2025), and 2 researchers independently included policies and extracted data. In this preliminary analysis, we report descriptive statistics stratified

by type of scholarly discipline and organization. The protocol and study registration can be found online.[1]

Results We screened 1008 organizations and included policies from 170 scholarly organizations: 100 journals, 20 publishers, 6 journal and publisher associations, 14 international multidisciplinary scholarly associations, 12 universities, and 18 funding agencies. From these organizations, we included 400 policies targeting different stakeholders (**Table 25-0987**): 159 policies (40%) relevant for researchers (eg, manuscript authors or grant applicants), 126 (32%) for peer reviewers, and 115 (29%) for editors or grant committee members. A total of 369 policies (92%) addressed both financial and other interests, 13 (3%) addressed exclusively financial interests, and 18 (5%) addressed exclusively other interests. A total of 350 policies (88%) addressed disclosure of interests and 287 (72%) addressed the management of conflicts of interest. Content describing both disclosure requirements and management of disclosed conflicts of interest could be found in 71 of 159 policies (45%) for researchers, 87 of 126 (69%) for peer reviewers, and 79 of 115 (69%) for editors or grant committee members. Enforcement strategies (eg, consequences of violating policies) to ensure the compliance with both disclosure and management requirements were addressed in 16 of 159 policies (10%) for researchers, 19 of 126 (15%) for peer reviewers, and 7 of 115 (6%) for editors or grant committee members.

Conclusions Most of the top scholarly organizations have conflict of interest policies addressing both financial and other types of interests. While almost all policies addressed disclosure of conflicts of interest, only slightly more than one-half addressed both disclosure of conflicts of interest and how disclosed conflicts of interest should be managed. Few policies described how they should be enforced.

Reference

1. Krejcikova HVB, Korfitsen CB, Bero L, et al. Conflicts of interest in research across scholarly disciplines: cross-sectional study and qualitative content analysis of policies. October 16, 2023. Accessed July 15, 2023. https://osf.io/vwn6c

[1]Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark, hkrejcikova@health.sdu.dk; [2]Open Patient Data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark; [3]Center for Bioethics and Humanities, University of Colorado Anschutz Medical Campus, Aurora, CO, US; [4]Yale School of Management, New Haven, CT, US; [5]North Carolina State University College of Veterinary Medicine, Raleigh, NC, US; [6]Lawrence Bloomberg Faculty of Nursing, University of Toronto, Toronto, Ontario, Canada; [7]School of Technology, Policy and Management, Delft University of Technology, Delft, the Netherlands; [8]Centre for Information and Innovation Law (CIIR), Faculty of Law, University of Copenhagen, Copenhagen, Denmark; [9]Department of Respiratory Medicine and Infectious Diseases, Copenhagen University Hospital–Bispebjerg and Frederiksberg, Copenhagen, Denmark.

Conflict of Interest Disclosures Lisa Bero is a meta-research section editor for *PLoS Biology*, a Conflict of Interest Advisor for Health Canada, and a Senior Research Integrity Editor at Cochrane, for which the University of Colorado receives remuneration. David C. Dorman is an associate editor for *Animals and Critical Reviews in Toxicology*. Ibo van de Poel is the integrity officer of the Delft University of Technology; sits on the editorial boards of *Science and Engineering Ethics*, *AI and Ethics*, and *International Journal of Technoethics*; and contributed to the book series *SpringerBriefs on Ethical and Legal Issues in Biomedicine and Technology*. Asbjørn Hróbjartsson is an associate editor of the *Journal of Evidence-Based Medicine* and the Cochrane Methodology Review Group and sits on the editorial board of the *Journal of Clinical Epidemiology*. Andreas Lundh sits on the editorial board of *BMC Medical Ethics*. Quinn Grundy is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

A Taxonomy-Based Guideline Framework for Conflict of Interest Disclosures (CoST)

Pritha Sarkar,[1] Ruth Whittam,[1] Leslie D. McIntosh[1]

Objective Properly declaring conflicts of interest (COIs) is essential for assessing potential bias in scholarly publications and maintaining research integrity. However, COI declarations are often inconsistent with significant variation

Table 25-0987. Number of Policies Targeting Selected Stakeholders Within Organizations and Scholarly Disciplines Included in This Study

Type of scholarly organization; stakeholders targeted by its policies	Type of scholarly discipline						
	Multidisciplinary	Economics	Engineering	Environmental toxicology and chemistry	Law	Medicine	Total
Journal; authors, peer reviewers, and editors	0	51	53	56	42	55	257
Publisher; authors, peer reviewers, and editors	49	0	3	0	0	4	56
Journal and publisher association; authors, peer reviewers, and editors	8	0	0	0	0	6	14
Multidisciplinary association; researchers and reviewers	23	0	0	0	0	0	23
University; researchers (academic staff), reviewers, and editors	15	0	0	0	0	0	15
Funding agencies; applicants, reviewers, and grant committee members	21	0	2	3	0	9	35
Total	116	51	58	59	42	74	400

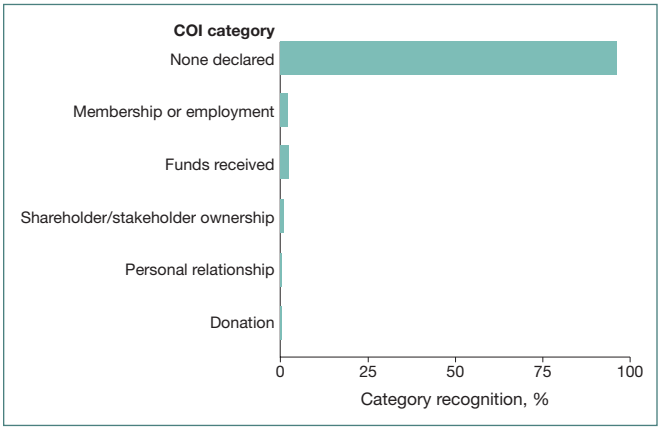
in the requirements set by different journals. This leads to discrepancies in reporting and contributes to confusion regarding what constitutes a complete and transparent COI disclosure. This study uses a data-driven approach to create and measure a COI declaration taxonomy akin to CReDiT.[1] If journals standardize COI reporting requirements to this taxonomy, it will help address these inconsistencies.

Design In 2023, Dimension's full-text database[2] was used to perform a descriptive study on COI statements found in 2966 randomly selected open-source research papers published between 2012 and 2023. COI declarations were classified in 6 primary categories: (1) none declared, (2) membership or employment, (3) stakeholder or shareholder ownership, (4) funds received, (5) personal relationship, and (6) donation. Using natural language processing tools, ~10% (297) of the statements from the initial pool of 2966 statements were manually annotated, creating a standard for automatic annotation of COI statements. Later, our analysis also leveraged the Retraction Watch[3] database to assess COI-related retractions when it became public in September 2023.

Results Our model, which identifies both undeclared COIs and its categories, found that only 896,893 of 16,044,369 gathered COI statements mentioned at least 1 COI category. The skewness among the recognized COI categories is depicted in **Figure 25-1072**. The data also revealed significant inconsistencies in COI disclosures (eg, variations in the structure and clarity of statements). Moreover, when the Retraction Watch database[3] was compared with the database for Dimensions,[2] it was observed that among 24,289 overlapping papers, 393 were retracted because of COIs. Of those 393 papers, 349 (~89%) had initially declared no COI.

Conclusions Due to the varying requirements imposed by different journals, authors often face confusion in clearly and concisely declaring COIs. As a result, many resort to minimal statements such as "none declared," and, in some instances, ambiguous terms such as "nil" or "none." This undermines the intent of transparency. The study highlights this critical gap in the standardization and understanding of COI declarations, emphasizing the need for a structured framework to enhance clarity and transparency. To bridge this gap, we propose Conflict of interestS Taxonomy (CoST), a guideline for COI reporting. This framework provides authors with a straightforward, questionnaire-style tool to ensure consistency and completeness in COI declarations. By fostering adherence to journal and institutional standards, this taxonomy may promote trust within the research community and strengthen the credibility of scientific publications. Our work highlights the importance of systemic change in COI reporting to safeguard the integrity of scholarly communication.

Figure 25-1072. Difference Among the Reported COI Categories



References

1. Contributor Role Taxonomy. Contributor Role Taxonomy (CRediT). 2022. https://credit.niso.org/

2. Hook DW, Porter SJ and Herzog C. Dimensions: Building Context for Search and Evaluation. *Front Res Metr Anal*. 2018;3:23. doi:10.3389/frma.2018.00023

3. The Retraction Watch Database. Retraction Watch Version 1.0.8.0. http://retractiondatabase.org/

[1]Digital Science, London, UK, p.sarkar@digital-science.com.

Conflict of Interest Disclosures All authors are employees of Digital Science. None of the authors received any donation or funds for the study.

Navigating the Challenges of Competing Interest Disclosures in Academic Publishing

Julia Gunn,[1] Coromoto Power Febres,[1] Laura Wilson[1]

Objective The disclosure of competing interests in published academic research is vital to transparency and the integrity of the scholarly record.[1] In recent years, Taylor & Francis has seen an increase in complex ethics cases across all disciplines involving perceived competing interests. This study examined this increase, identified patterns across cases, and considered the future implications for academic publishing.

Design This qualitative analysis examined all competing interest cases (N = 71) investigated by the Taylor & Francis Publishing Ethics & Integrity (PEI) team between 2021 and 2024. The study began with data from 2021 because that is when the PEI team began systematically classifying cases. These cases included both unpublished and published manuscripts in 62 journals across all subject areas and included cases involving authors, editors, and peer reviewers. The years in which cases were raised are not necessarily the same as the date published. The study examined case volume increases over time to highlight trends by discipline and subject area, case complexity, and competing interest type (financial or nonfinancial). It also analyzed year-on-year (YoY) growth to identify whether there had been sudden growth in any 1-year period. Further analysis was conducted of the available data for the year that displayed sudden growth (2024). The study sought to explain the increase and consider ways publishers can refine policies to adequately address growing concerns about perceived competing interests.

Results This study revealed several trends. First, between 2021 and 2023, the number of cases increased annually

(2021: 9; 2022: 12; 2023: 14; 2024: 35), with significant YoY growth between 2023 and 2024, marking a 150% increase from the previous year. Second, the study also revealed an increase in allegations of failure to disclose perceived nonfinancial competing interests, such as personal relationships, ties to political or ideological organizations, and other relationships perceived to be relevant to the published content. In 2024, the majority of cases involved nonfinancial competing interests (financial: 6; nonfinancial: 19; financial and nonfinancial: 4), with 6 excluded due to insufficient data. Of the 71 cases, 3 involved allegations against editors, 2 were against reviewers, and the remaining 66 concerned authors.

Conclusions This study highlights the need for greater consensus regarding competing interest disclosure. Although publishers typically require disclosure of financial and nonfinancial competing interests,[2] this study suggested that authors, editors, and publishers may struggle to determine what requires disclosure. Limitations of this study include its small sample size and focus only on cases investigated by the ethics team. Additionally, the study suggested that researchers may benefit from education on perceived competing interests. Further research is needed to explore the broader landscape of competing interest disclosure across disciplines and publishers to enhance the transparency of the scholarly record.

References

1. Gibson DS, O'Hanlon R. The current conflict of interest landscape and the potential role of information professionals in supporting research integrity. *J Hosp Libr*. 2020;20(3):183-203. doi:10.1080/15323269.2020.1778970

2. Resnik D. Disclosing and managing non-financial conflicts of interest in scientific publications. *Res Ethics*. 2023;19(2):121-138. doi:10.1177/17470161221148387

[1]Taylor & Francis Group, Philadelphia, PA, US, julia.gunn@taylorandfrancis.com.

Conflict of Interest Disclosures Julia Gunn is a full-time employee of Taylor & Francis. Coromoto Power Febres is a full-time employee of Taylor & Francis and a member of the International Association of Scientific, Technical & Medical Publishers (STM) Research Integrity Committee and United2Act Education and Awareness Working Group. Laura Wilson is a full-time employee of Taylor & Francis and a member of the STM Membership Committee, United2Act Trust Markers Working Group, and Association of Learned and Professional Society Publishers Policy Committee.

Virtual

Conflict of Interest Network Robustness and Funder Homogeneity Associated With Reported Adverse Events and Deaths in Published Drug Studies

S. Scott Graham,[1] Joshua B. Barbour,[2] Zoltan P. Majdik,[3] Madeline Bruegger,[1] Carlee A. Baker,[1] Justin F. Rousseau[4]

Objective Research on industry funding and author conflicts of interest (COIs) suggests that centralization of funding may lead to detrimental effects on biomedical research.[1,2] This study was conducted to assess relationships between COI networks and drug adverse event rates reported to the US Food and Drug Administration.

Design We used a GPT 4.0 pipeline (a series of computational steps applied to a dataset for transformation and/or analysis) to identify and classify reported author COIs and funders associated with 302 of the most commonly used drugs in the US. Included articles included up to 150 of the most relevant PubMed results when searching for each drug's proprietary and generic name. For each declared COI, the funder was identified and funder types were classified by a fine-tuned GPT 4.0 model as industry, federal, university, patient advocacy organization, professional society, university, or hospital/clinic. COI relationships for each drug product were assembled into network models mapping relationships among all reported funders and all articles for each product. Each network was evaluated for network robustness by measuring critical failure threshold (CFT) (ie, the minimum number of nodes that would need to be removed to result in a 50% collapse of the network).[3] COI network robustness is a useful measure of the centralization of the funding network as a low number of funder networks would have lower failure thresholds. A network with many independent funders would require more nodes to be removed before network failure. Controlling for COI rates, drug features (over-the-counter availability, generic availability, US Drug Enforcement Administration schedule, usage rates), and publication features (mean article year of publication, total authorships in the network), we evaluated the relationship between network robustness and reported adverse events rates (total, serious, and fatal). A secondary model was fit to determine whether individual funder type frequencies or heterogeneity were associated with failure threshold. The GPT classifier was compared with a human annotated sample and evaluated for precent accuracy.

Results We identified 37,993 articles associated with the 302 drugs. The GPT 4.0 classifier achieved 98.9% accuracy compared with the human annotated sample. In the adverse events models, CFT returned adjusted incident rate ratios (IRRs) of 0.94 (95% CI, 0.91-0.97; $P < .01$) for total events, 0.94 (95% CI, 0.91-0.97, $P < .01$) for serious events, and 0.94 (95% CI, 0.91-0.98; $P < .01$) for fatal events. The number of funders per network was not a significant factor in any adverse events model. The network features model (**Table 25-0837**) returned significant relationships between the CFT and the number of funders in the top 10% by out-degree, a network science measure of the number of outgoing connections from a given node (IRR, 1.03 [95% CI, 1.02-1.04]; $P < .001$) and funder-type heterogeneity (IRR, 1.19 [95% CI, 1.12-1.26]; $P < .001$).

Conclusions In this study, COI network robustness was inversely associated with adverse events, including serious and fatal events, reported in drug studies. Secondary analyses indicate that this was associated with funding network centralization and homogeneity. The products evaluated in

Table 25-0837. Generalized Linear Model Results for Relationships Between Network Features and Critical Failure Threshold (N = 301)[a]

Factor	Median (IQR)	Range	Incident rate ratio (95% CI)	*P* value
Intercept			1.37 (0.07-25.14)	.83
Funder type, %				
Industry	0.91 (0.13)	0.31-1.00	2.65 (0.15-48.38)	.51
Federal	0.03 (0.06)	0-0.48	0.49 (0.1-17.33)	.67
Patient advocacy organization	0.00 (0.01)	0-0.31	0.28 (0.01-35.14)	.58
Professional society	0.01 (0.02)	0-0.4	0.53 (0.01-26.17)	.75
University	0.01 (0.02)	0-0.25	0.81 (0.01-110.07)	.93
Hospital/clinic	0.00 (0.01)	0.01-0.40	0.86 (0.01-94.48)	.95
No. of top 10% funders	0.08 (0.02)	0-0.44	1.03 (1.02-1.04)	<.001[b]
No. of funder types	5.00 (2.00)	1-6	1.19 (1.12-1.26)	<.001[b]

[a]Nagelkerke $R^2 = 0.765$.

[b]Indicates statistical significance.

this study were associated with lower adverse event rates when the collected research had a larger number of funders, regardless of funder type. This suggests that research funding policy should encourage funding diversity.

References

1. Graham SS, Harrison KR, Edward JC, Majdik ZP, Barbour JB, Rousseau JF. Beyond bias: aggregate approaches to conflicts of interest research and policy in biomedical research. *World Med Health Policy*. 2024;16(3):489-505. doi:10.1002/wmh3.608.

2. Sismondo S. Ghost management: how much of the medical literature is shaped behind the scenes by the pharmaceutical industry? *PLoS Med*. 2007;4(9):e286. doi:10.1371/journal.pmed.0040286

3. Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. *Nature*. 2000;406(6794):378-382. doi:10.1038/35019019

[1]Department of Rhetoric & Writing, The University of Texas at Austin, Austin, TX, US, ssg@utexas.edu; [2]Department of Communication, University of Illinois Urbana-Champaign, Urbana, IL, US; [3]Department of Communication, North Dakota State University, Fargo, ND, US; [4]Department of Neurology & Peter O'Donnell Jr. Brain Institute, The University of Texas Southwestern Medical Center, Dallas, TX, US.

Conflict of Interest Disclosures S. Scott Graham received grant funding from the National Institute of General Medical Sciences (NIGMS). Joshua B. Barbour received grant funding from the NIGMS, NSF, IC2, and Blue Cross Blue Shield; royalties form Cengage Publishing; and consulting fees from The Aslan Group and BrainCheck. Justin F. Rousseau received grant funding from the NIGMS, National Library of Medicine, National Institute on Aging, National Institute of Allergy and Infectious Diseases, National Institute of Mental Health, National Center for Advancing Translational Sciences, Texas Child Mental Health Care Consortium, Michal & Susan Dell Foundation, Texas Alzheimer's Research and Care Consortium, Austin Public Health, and the Health Care Cost Institute. No other disclosures were reported.

Funding/Support Funding for this research was provided by the NIGMS of the National Institutes of Health under award number R01GM141476.

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 abstract; and decision to submit the abstract for presentation.

Disclosed and Undisclosed Conflicts of Interest in US Guidelines for the Management of Obesity

Alessandro Bianconi,[1] Matteo Fiore,[1] Maria Elena Flacco,[2] Lamberto Manzoli[1]

Objective Despite global initiatives to improve transparency regarding conflicts of interest (COIs) in scientific publications, the disclosure of competing interests remains inadequate among authors of clinical practice guidelines (CPGs), as well as for other types of publications.[1,2] Open access databases like Open Payments offer a potential resource for assessing undisclosed COIs in scientific literature.[3] This study evaluates the prevalence of undisclosed industry-related financial COIs among authors of US-issued CPGs on obesity management from 2016 to present.

Design On June 3, 2024, we identified CPGs on obesity management issued by US agencies through 3 repositories (UpToDate, Medscape, and Guidelines International Network). COI statements for each guideline were extracted. To identify undisclosed industry-related financial COIs, we cross-referenced authors who reported no competing interests with the Open Payments database,[3] considering only any form of payment and fund received prior to the guideline publication date. Undisclosed COIs were classified based on cumulative amounts as greater than or equal to $10,000 and less than $10,000. The presence of COIs was summarized as absolute and relative frequencies. Given the purely descriptive nature of this study, no hypotheses were tested.

Results We retrieved 5 CPGs (American Academy of Pediatrics 2023, American Diabetes Association 2023, American Gastroenterological Association 2022, Veteran Affairs/Department of Defense 2020, Association of American Clinical Endocrinologists/American College of Endocrinology 2016) (**Table 25-1155**). Among 145 authors, 80 (55.2%) received payments from industry agents related to obesity prevention or treatment. Of these, 33 authors (41.3% [22.8% overall]) had undisclosed COIs. Among undisclosed COIs, 11 (33.3%) were greater than or equal to $10,000 and 22 (66.7%) were less than $10,000. Every guideline had at least 1 chair or first author with industry-related COIs. Seven of 9 (77.8%) of these prominent authors did not disclose any COI, though all undisclosed COIs were less than $10,000.

Conclusions We found a relevant amount of undisclosed COIs in obesity management CPGs, and this amount may be underestimated because only financial interests could be evaluated. The presence of undisclosed COIs among CPG authors may reflect the complexities of COI disclosure rather than a straightforward lack of transparency. Variations in COI policies across scientific societies and publishers, which

Table 25-1155. US Guidelines on Obesity Management Published From 2016 to 2024

Guideline	No. of authors with COI/ total No.	No. of authors with undisclosed COI ≥$10,000/total No. of undisclosed COIs	No. of chairs with COI/total No. of chairs	No. of chairs with undisclosed COI ≥$10,000/total No. undisclosed COIs
American Academy of Pediatrics 2023	8/21	1/8	1/1	0/1
American Diabetes Association 2023	26/51	6/10	1/1	0/1
American Gastroenterological Association 2022	7/10	2/7	2/2	0/2
Veterans Affairs/Department of Defense 2020	4/36	1/4	3/4	0/3
Association of American Clinical Endocrinologists/American College of Endocrinology 2016	26/27	1/4	1/1	0/1
Total	80/145	11/33	8/9	0/8

Abbreviation: COI, conflicts of interest.

indicate the amount and type of payments that represent a COI that should be disclosed, may result in inconsistent reporting standards. This inconsistency presents an opportunity to harmonize COI management policies across publishing entities, promoting more comprehensive disclosure practices. Implementing regular monitoring of COI statements by peer reviewers and editors can further enhance transparency and trustworthiness. Tools like Open Payments can aid this process, though they have limitations, including potential inaccuracies and exclusion of nonfinancial COIs. Addressing these issues can strengthen the reliability of CPGs and other scientific publications. Further research should investigate the prevalence of undisclosed and overall COIs in other fields, assessing their potential impact on guideline recommendations and, thus, on real-world clinical practice.

References

1. Khan R, Scaffidi MA, Rumman A, Grindal AW, Plener IS, Grover SC. Prevalence of financial conflicts of interest among authors of clinical guidelines related to high-revenue medications. *JAMA Intern Med*. 2018;178(12):1712-1715. doi:10.1001/jamainternmed.2018.5106

2. Lenzer J, Hoffman JR, Furberg CD, Ioannidis JPA. Ensuring the integrity of clinical practice guidelines: a tool for protecting patients. *BMJ*. 2013;347:f5535. doi:10.1136/bmj.f5535

3. Open Payments. Accessed February 14, 2025. https://openpaymentsdata.cms.gov

[1]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy, alessandro.bianconi4@studio.unibo.it; [2]Department of Environmental and Prevention Sciences, University of Ferrara, Ferrara, Italy.

Conflict of Interest Disclosures None reported.

Data Sharing and Access

In-person

Data Availability Statements in Health Research in Articles and Journal Policies in Korea

Sue Kim,[1] Soo Young Kim,[2] Hyun Jung Yi[3]

Objective Data sharing is strongly recommended for replicability, trustworthy findings, and generation of new hypotheses.[1] Although a data availability statement (DAS) reveals whether data are shared and how to access data, little is known on DAS patterns. This study aimed to evaluate the proportion of articles presenting a DAS according to journal policies in Korea.

Design This cross-sectional study was conducted from July 2024 to January 2025. Articles in KoreaMed member journals published in 2023 were randomly selected (5%) by generating random numbers using the RANDBETWEEN function in Microsoft Excel 2021. The inclusion criterion was human studies in Korean or English. Narrative reviews, letters, editorials, or case studies were excluded. Selected articles (600 of 12,000) were evaluated by the following parameters: presence of a DAS, its appropriateness, and whether the journal has a DAS policy statement. The research design, adherence to the International Committee of Medical Journal Editors (ICMJE) clinical trials data sharing requirement,[2] mention of reporting guidelines in the journal's guidelines, and health discipline were also assessed as related factors. Two researchers independently reviewed the full text and discussed for consensus, using the Design Algorithm for Medical Literature on Intervention tool[3] to determine study design.

Results After excluding irrelevant articles, 243 articles (including surveys, cohort studies, randomized clinical trials [RCTs], systematic reviews, and integrative reviews) were identified from 126 journals; 74 journals (58.7%) had a DAS policy and 155 articles (63.8%) were published in such journals. Of these, however, 96 articles (61.9%) did not comply with the journal's policy of specifying data sharing. Of the 243 articles, only 67 (27.6%) articles (from 35 journals) presented a DAS, of which 19 (28.4%) were inappropriate, 41 (61.2%) were generic cliché statements, and only 4 (6%) provided a hyperlink to the data. Of the 74 journals with a DAS policy, 7 (9.5%) presented mixed results ranging from inappropriate to satisfactory DAS, suggesting the need for editorial consistency. Of the 16 experimental studies (from 13 journals), only 6 (3 RCTs, 3 non-RCTs) complied with ICMJE's policy requiring a DAS, of which only 1 linked to a data registry. For observational studies, 58 of 216 (26.8%) presented a DAS. Journal policy on reporting guidelines (52 articles) was not significant for DAS or health discipline, but variation in the proportion of articles with an acceptable DAS

was observable: 66% of articles from medicine (31), 67% from dentistry (2), 100% from nursing (9), and 75% from public health or nutrition (6).

Conclusions Although roughly 62% of journals had a DAS policy, less than half of articles presented a DAS, with only 6% providing actual data. Given this gap, greater awareness of DASs among researchers and editorial consistency is needed, especially considering that ICMJE policy requires a DAS in clinical trials involving humans.

References

1. Hulsen T. Sharing is caring-data sharing initiatives in healthcare. *Int J Environ Res Public Health*. 2020;17(9):3046. doi:10.3390/ijerph17093046

2. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *JAMA*. 2017;317(24):2491-2492. doi:10.1001/jama.2017.6514

3. Seo HJ, Kim SY, Lee YJ, et al. A newly developed tool for classifying study designs in systematic reviews of interventions and exposures showed substantial reliability and validity. *J Clin Epidemiol*. 2016;70:200-205. doi:10.1016/j.jclinepi.2015.09.013

[1]College of Nursing, Mo-Im Kim Nursing Research Institute, Yonsei University, Seoul, South Korea, suekim@yuhs.ac; [2]Department of Family Medicine, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul, South Korea; [3]Medical Library, Hanyang University Guri Hospital, Guri, South Korea.

Conflict of Interest Disclosures None reported.

Funding/Support This study was supported by the Korean Council of Science Editors.

Role of the Funder/Sponsor The funder supported the costs of the research and was not involved in the planning, analysis, or interpretation of the results of the study.

Metrics of Primary and Secondary Publications of Clinical Trials With Data Shared on the YODA Project Platform

Erfan Taherifard,[1] Hollin R. Hakimian,[1] Maryam Mooghali,[1] Sahil R. Mane,[1] Mengyuan Fu,[1] Stephen Bamford,[2] Karla Childers,[3] Nihar R. Desai,[4] Cary P. Gross,[1,5,6] Debbie Hewens,[2] Harlan M. Krumholz,[4,7,8] Richard Lehman,[9] Jessica D. Ritchie,[7] Tamsin Sargood,[2] Joshua D. Wallach,[10] Molly K. Willeford,[7] Joseph S. Ross[1,7,8]

Objective Sharing clinical trial data can advance scientific knowledge, maximize research impact, and reduce duplication, costs, and risks.[1] However, the impact of such initiatives has not been systematically evaluated. The Yale Open Data Access (YODA) Project is an academic-based data-sharing initiative housed at Yale University.[2] This study evaluated the impact of studies published using data from clinical trials sponsored by Johnson & Johnson that are shared on the YODA Project platform, distinguishing between studies led by the primary study investigators and external investigators.

Design We conducted a cross-sectional study of clinical trials sponsored by Johnson & Johnson that were shared on the YODA Project platform as of December 31, 2021; trials without a primary publication were excluded. Primary publications—defined as the first full-length peer-reviewed published article reporting the trial's primary end point—were identified. Secondary publications—defined as peer-reviewed published studies using individual-level participant data—were retrieved by screening citations to each trial's primary publication in the Web of Science. The primary searches in the Web of Science were conducted from September 15 to December 15, 2024, with all searches updated as of December 15, 2024. We collected metrics for primary and secondary publications, including Impact Factors of publishing journals and annual citation counts, Altmetric scores, and Mendeley reader counts. Annual metrics were calculated by dividing each metric by the number of days since publication, then multiplying by 365. Secondary publications were categorized as internal (authored by primary study investigators) or external. Comparisons were performed using Mann-Whitney *U* tests.

Results There were 294 trials with a primary publication sharing data on the YODA Project platform, 218 (74.1%) of which had at least 1 secondary publication, totaling 1699 secondary publications. Most were pooled analyses (1302 [76.6%]), as opposed to secondary analyses (397 [22.4%]). Trials of heart and blood disease interventions had the highest number of secondary publications per trial (median [IQR], 10.5 [2-19] publications). Compared with primary publications, secondary publications were published in journals with lower median (IQR) journal Impact Factors (8.4 [4.5-22.5] vs 4.9 [3.0-8.5]) and had lower annual citation counts (15.9 [7.8-45.4] vs 5.1 [2.5-11.4]), Altmetric scores (0.8 [0.3-3.1] vs 0.6 [0.2-3.1]), and Mendeley reader counts (10.1 [5.1-17.9] vs 8.4 [3.3-15.5]) (**Table 25-1008**). Secondary publications were predominantly internal (1169 [68.8%]), which, when compared with external secondary publications, were published in journals with higher median (IQR) Impact Factors (5.1 [3.2-10.3] vs 4.5 [2.9-7.0]) and had higher annual citation counts (5.8 [2.7-12.6] vs 3.2 [1.9-7.5]), Altmetric scores (0.7 [0.2-2.8] vs 0.5 [0.1-3.8]), and Mendeley reader counts (8.9 [4.3-16.3] vs 6.9 [0.9-13.1]).

Conclusions Clinical trial data sharing through the YODA Project has fostered substantial scholarship, with nearly one-third of secondary studies published using these trials' data by external investigators not affiliated with the primary study team. While secondary publications generally had lower impact metrics than primary publications, they still demonstrated meaningful scientific dissemination and usage, underscoring their contributions to the broader medical research enterprise.

References

1. Angraal S, Ross JS, Dhruva SS, Desai NR, Welsh JW, Krumholz HM. Merits of data sharing: the Digitalis Investigation Group trial. *J Am Coll Cardiol*. 2017;70(14):1825-1827. doi:10.1016/j.jacc.2017.07.786

Table 25-1008. Primary and Secondary Publication Impact Metrics for Studies Using Data From Clinical Trials Sponsored by Johnson & Johnson Shared on the YODA Project Platform

Metric	All publications (N = 1993)	Primary publications (n = 294)	Secondary publications (n = 1699)	*P* value[a]	Secondary publications		
					Internal (n = 1169)	External (n = 530)	*P* value[a]
Journal Impact Factor, median (IQR)	5.1 (3.2-10.1)	8.4 (4.5-22.5)	4.9 (3.0-8.5)	<.001	5.1 (3.2-10.3)	4.5 (2.9-7.0)	<.001
Annual citation count, median (IQR), No.	5.9 (2.6-13.4)	15.9 (7.8-45.4)	5.1 (2.5-11.4)	<.001	5.8 (2.7-12.6)	3.2 (1.9-7.5)	<.001
Annual Altmetric score, median (IQR)	0.7 (0.2-3.1)	0.8 (0.3-3.1)	0.6 (0.2-3.1)	.004	0.7 (0.2-2.8)	0.5 (0.1-3.8)	.06
Annual Mendeley reader count, median (IQR), No.	8.6 (3.7-15.8)	10.1 (5.1-17.9)	8.4 (3.3-15.5)	<.001	8.9 (4.3-16.3)	6.9 (0.9-13.1)	<.001

[a]*P* values were calculated using the Mann-Whitney *U* test for comparisons between groups.

2. 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;5:180268. doi:10.1038/sdata.2018.268

[1]Section of General Internal Medicine, Yale School of Medicine, New Haven, CT, US, erfan.taherifard@yale.edu; [2]Johnson & Johnson, London, England, UK; [3]Johnson & Johnson, New Brunswick, NJ, US; [4]Section of Cardiovascular Medicine, Yale School of Medicine, New Haven, CT, US; [5]Cancer Outcomes, Public Policy, and Effectiveness Research Center, Yale School of Medicine, New Haven, CT, US; [6]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, US; [7]Yale–New Haven Hospital Center for Outcomes Research and Evaluation, New Haven, CT, US; [8]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, US; [9]Institute of Applied Health Research, University of Birmingham, Birmingham, England, UK; [10]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, US.

Conflict of Interest Disclosures Erfan Taherifard is an academic editor of *PLOS One*. Stephen Bamford, Karla Childers, Debbie Hewens, and Tamsin Sargood are current employees and shareholders of Johnson & Johnson. Nihar R. Desai, Cary P. Gross, Harlan M. Krumholz, Jessica D. Ritchie, Joshua D. Wallach, Molly K. Willeford, and Joseph S. Ross are members of the YODA Project leadership team and receive support from Johnson & Johnson to support data sharing through the YODA Project. Nihar R. Desai reported working under contract with the Centers for Medicare and Medicaid Services to develop and maintain performance measures used for public reporting and pay-for-performance programs and receiving research grants from and consulting for Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, CSL Behring, Cytokinetics, Merck, Novartis, scPharmaceuticals, and Vifor. Cary P. Gross has received research funding from the NCCN Foundation (AstraZeneca) and Genentech and is an associate editor of *JAMA Internal Medicine*. Harlan M. Krumholz reported receiving options for Element Science and Identifeye and payments from F-Prime for advisory roles, being a cofounder of and holding equity in Hugo Health, Refactor Health, and ENSIGHT-AI, being associated with research contracts through Yale University from Janssen, Kenvue, and Pfizer, and being editor in chief of *JACC*. Jessica D. Ritchie reported receiving support from the US Food and Drug Administration (FDA) and Arnold Ventures. Joshua D. Wallach reported receiving support from the FDA and Arnold Ventures, having previously served as a consultant to Hagens Berman Sobol Shapiro LLP and Dugan Law Firm APLC, and being an associate editor of *JACC*. Joseph S. Ross reported receiving support from the FDA, Arnold Ventures, Agency for Healthcare Research and Quality, and National Heart, Lung, and Blood Institute and being 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 that was settled in September 2022, and being a deputy editor of *JAMA*.

Funding/Support This work was supported in part by a research grant from Johnson & Johnson through Yale University to establish a platform for clinical trial data sharing.

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; or decision to submit the abstract for presentation.

Data Sharing Statement Reporting Across Medical Specialties

Eli Paul,[1] Griffin Hughes,[1] Alex Hagood,[1] Matt Vassar[1,2]

Objective Data sharing promotes secondary use, transparency, and reproducibility, yet prior studies have revealed gaps in policy standardization and implementation.[1] Although some evaluations have explored data sharing statement (DSS) practices within specific fields or policy contexts, no broad, comparative assessment across medical specialties exists.[2,3] This meta-analysis assesses DSS prevalence across 19 medical specialties and examines variability in adoption rates.

Design We conducted a meta-analysis using harmonized article-level data from 19 separate studies, each prospectively designed and executed by our research team using the same methodology, with each study focused on a distinct medical specialty. Between December 2023 and July 2024, all studies systematically searched PubMed for original research articles published between 2018 and 2023. Study designs included in our study were clinical trials, cohort studies, cross-sectional studies, case series, case-control studies, cost-effectiveness analyses, qualitative research, survey-based studies, and other designs that involved human participants. Reviews, editorials, and commentaries were excluded. For each article, we assessed the presence of a DSS and whether data were publicly accessible. Two independent reviewers performed screening and data extraction per specialty, with discrepancies resolved by consensus. We conducted a random-effects meta-analysis using inverse-variance weights based on standard errors of DSS proportions. Influential study diagnostics were applied to identify and mitigate the impact of extreme specialty effects. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guideline was followed for the design and execution of our study.

Results Among 20,095 eligible articles, 6092 (30.3%) reported a DSS (**Figure 25-1134**). Prevalence varied substantially by specialty, from 63.9% in Rheumatology to 1.5% in Plastic Surgery. The pooled DSS reporting rate was 29.8% (prediction interval, 0-65.9%), with high heterogeneity (I^2 = 99.60%). After removing influential specialties (rheumatology, neurology, and plastic surgery), the adjusted pooled rate was 27.5% (prediction interval, 0-54.9%), with high heterogeneity persisting (I^2 = 99.14%).

Conclusions Despite increasing emphasis on open science, DSS reporting remains low and inconsistent across medical specialties. The high heterogeneity suggests that DSS reporting is shaped more by specialty-specific norms and editorial practices than by standardized policy. By conducting 19 internally coordinated studies using identical methodology, our team generated a uniquely standardized dataset that enabled robust, cross-specialty comparisons. Stronger policy enforcement and cross-disciplinary coordination among journals, funders, and regulatory bodies are needed to improve transparency and foster responsible data sharing in medical research.

References

1. Johnson AL, Anderson JM, Bouvette M, et al. Clinical trial data-sharing policies among journals, funding agencies, foundations, and other professional organizations: a scoping review. *J Clin Epidemiol*. 2023;154:42-55. doi:10.1016/j.jclinepi.2022.11.009

2. Tedersoo L, Küngas R, Oras E, et al. Data sharing practices and data availability upon request differ across scientific disciplines. *Sci Data*. 2021;8:192. doi:10.1038/s41597-021-00981-0

3. Vassar M, Jellison S, Wendelbo H, Wayant C. Data sharing practices in randomized trials of addiction interventions. *Addict Behav*. 2020;102:106193. doi:10.1016/j.addbeh.2019.106193

[1]Office of Medical Student Research, Oklahoma State University Center for Health Sciences, Tulsa, OK, US, eli.paul@okstate.edu; [2]Department of Psychiatry and Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, OK, US.

Conflict of Interest Disclosures Matt Vassar reported receiving funding from the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism, the US Office of Research Integrity, and the Oklahoma Center for Advancement of Science and Technology and internal grants from the Oklahoma State University Center for Health Sciences outside the submitted work. No other disclosures were reported.

Diversity and Inclusion

In-person

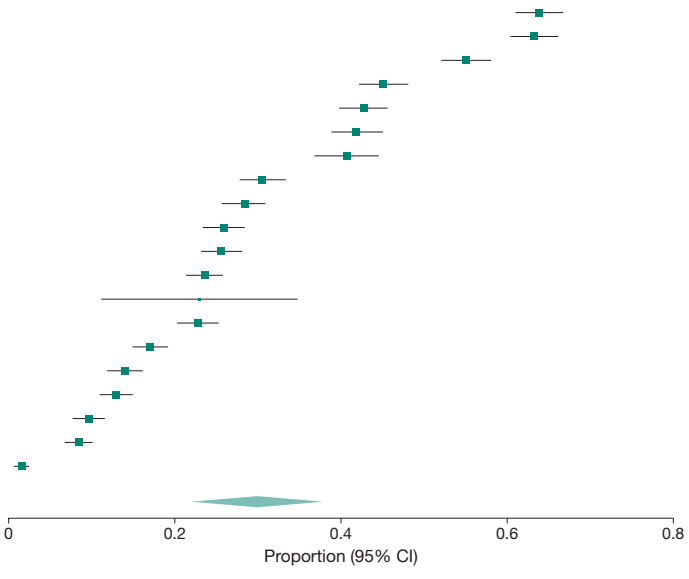
Retraction Prevalence and Gender Imbalance Among Highly Cited Authors and All Authors Across Scientific Disciplines

John P. A. Ioannidis,[1,2,3,4] Angelo Maria Pezzullo,[4,5] Antonio Cristiano,[4,5] Guillaume Roberge,[6] Stefania Boccia,[5,7] Jeroen Baas[8]

Objective Although retractions are increasingly frequent, they remain a small fraction of publications. We have previously incorporated retraction data into Scopus-based databases of top-cited (top 2%) scientists to facilitate linkage of retractions with impact metrics at the individual scientist level.[1] Here, we explored whether gender disparities in the

Figure 25-1134. Pooled Rates of Data Sharing Statement Publication in Medical Journal Articles

Medical specialty	No. of articles with a data sharing statement: Yes	No	Total No. of articles	Proportion (95% CI)
Rheumatology	681	384	1065	0.64 (0.61-0.67)
Neurology	702	406	1108	0.63 (0.61-0.66)
Pediatrics	584	474	1058	0.55 (0.52-0.58)
Pulmonology	487	590	1077	0.45 (0.42-0.48)
Cardiology	477	639	1116	0.43 (0.40-0.46)
Gastroenterology	400	553	953	0.42 (0.39-0.45)
Pain management	245	357	602	0.41 (0.37-0.45)
Psychiatry	319	723	1042	0.31 (0.28-0.33)
Obstetrics and gynecology	319	806	1125	0.28 (0.26-0.31)
Dermatology	297	851	1148	0.26 (0.23-0.28)
Physical medicine and rehabilitation	327	951	1278	0.26 (0.23-0.28)
Ophthalmology	326	1059	1385	0.24 (0.21-0.26)
Orthopedic surgery and rheumatology	11	37	48	0.23 (0.11-0.35)
Addiction medicine	240	815	1055	0.23 (0.20-0.25)
Otolaryngology	201	979	1180	0.17 (0.15-0.19)
Orthopedic surgery	145	891	1036	0.14 (0.12-0.16)
General surgery	141	953	1094	0.13 (0.11-0.15)
Emergency medicine	84	791	875	0.10 (0.08-0.12)
Anesthesiology	95	1028	1123	0.08 (0.07-0.10)
Plastic surgery	11	716	727	0.02 (0.01-0.02)
Restricted maximum likelihood model (Q = 5312.05; df = 19; P < .001, I^2 = 99.6%; τ^2 = 0.03)				**0.30 (0.22-0.38)**



The size of each square represents the weight assigned to each specialty; the diamond represents the overall estimated pooled rate.

likelihood of having retractions exist, both among highly cited authors and among all authors with at least 5 publications.

Design On August 15, 2024, we screened 55,237 Retraction Watch records, excluding nonretractions, those clearly unrelated to author error, those tied to republished papers, or those not linkable to Scopus, leaving 39,468 eligible retractions. We examined demographics of scientists with and without retractions among highly cited authors (career-long: n = 217,097) and among all authors with at least 5 publications (n = 10,361,367). We were able to assign gender using NamSor[2] to 186,466 and 8,267,888 authors, respectively. We stratified authors by publication age, field,[3] country income level (high, other), and publication volume, identifying for all these strata and for individual countries, men and women, and with and without retractions. We computed gender-specific retraction rates and calculated the relative propensity (R) of women vs men to have at least 1 retraction.

Results Authors with retractions were more common among highly cited scientists (3.3%) than among non–highly cited scientists (0.7%). Overall, gender differences were modest: among highly cited authors, retraction rates were 2.9% for women and 3.1% for men; among all authors, retraction rates were 0.7% for both genders. Men consistently showed slightly higher retraction rates than women within both income groups. Field-specific analysis among all authors revealed women's rates were at least one-third lower than men's (R < 0.67) in biology, biomedical research, and psychology and cognitive sciences, but higher (R > 1.33) in economics and business, engineering, and information and communication technology. Among highly cited scientists, the highest women to men retraction ratios were in mathematics and statistics (R = 3.06) and engineering (R = 1.78), while biomedical research (R = 0.64) and built environment and design (R = 0.65) had lower rates for women. Across publication age cohorts, gender differences in retraction rates among all authors were minimal; however, among highly cited authors, younger cohorts showed increasingly higher rates among men (4.2% of men and 3.0% of women in those starting to publish in 2002-2011; 8.7% of men and 4.9% of women in those starting to publish post-2011). Country-level data revealed particularly large gender gaps in Pakistan (men, 28.7%; women, 14.3%), Iran (12.4% vs 9.3%), and India (9.2% vs 6.6%) among highly cited authors. Among all authors, country-level gender gaps were small.

Conclusions Gender differences in retraction rates were small in most settings but varied by field, country, and publication cohort. Overall, field and country were more strongly associated with retraction rates than gender. These results highlight the need to account for structural and contextual factors when interpreting gender disparities.

References

1. Ioannidis JPA, Pezzullo AM, Cristiano A, Boccia S, Baas J. Linking citation and retraction data reveals the demographics of scientific retractions among highly cited authors. *PLoS Biol*. 2025;23(1):e3002999. doi:10.1371/journal.pbio.3002999

2. NamSor. Accessed July 14, 2025. https://NamSor.app

3. Archambault É, Beauchesne OH, Caruso J. Towards a multilingual, comprehensive and open scientific journal ontology. In: *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics*. 2011;13:66-77.

[1]Department of Medicine, Stanford University, Stanford, CA, US, jioannid@stanford.edu; [2]Department of Epidemiology & Population Health, Stanford University, Stanford, CA, US; [3]Department of Biomedical Data Science, Stanford University, Stanford, CA, US; [4]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, US; [5]Section of Hygiene, Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Rome, Italy; [6]Analytics and Data Services, Elsevier B.V., Montreal, Canada; [7]Department of Women, Children and Public Health Sciences, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy; [8]Research Intelligence, Elsevier B.V., Amsterdam, the Netherlands.

Conflict of Interest Disclosures Guillaume Roberge and Jeroen Baas are employees of Elsevier. John P. A. Ioannidis is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Additional Information Elsevier runs Scopus, which is the source of these data, and also runs the repository where the database of highly cited scientists is now stored.

Academic Institutional Affiliations and Gender of Authors, Editorial Board Members, and Editors of Journals

Ulrike K. Müller,[1] M. Janneke Schwaner,[2] Ksenia Keplinger[2]

Objective Despite decades of diversity initiatives, inequality persists, including in academic publishing. Promising trends remain weak when intersectionality is considered.[1,2] Inspired by the concept of intersectionality, we hypothesize that relying solely on gender as an indicator underestimates inequality, as decades of tokenistic inclusion measures have masked the true extent of inequality. The central idea of this study requires us to go beyond direct equity indicators, such as gender, race, and geographic location, and to identify indirect indicators. We predict that indirect indicators of inequality that correlate with gender due to intersectionality (eg, type of institution, as the proportion of women faculty is lower at research-intensive institutions[3]) will show higher levels of inequality, with more extreme disparities at higher rungs of the academic power ladder.

Design In this cross-sectional study, we quantified inequality predicted by gender vs type of institution by collecting data from 50 top-ranked journals (SCImago ranking) for 6 academic fields (**Table 25-1014**) on lead editors, on editorial boards for 5 of those journals, on corresponding authors for 1 journal, and the editorial board for 1 journal from 1979 to 2022. All data were accessed in 2022. We determined gender from pronouns used in self-authored or self-edited documents whenever possible, such as editorial biographical

Table 25-1014. Number and Proportion of Women and Scholars

Position	Total, No.	No. (%)	
		Women	PUI
Lead editor			
Zoology	67	16 (24)	1 (1)
Cell biology	62	24 (39)	0
Molecular biology	60	23 (38)	0
Ecology and evolution	78	29 (37)	1 (1)
General	63	13 (21)	0
Education	77	35 (45)	2 (3)
Median, %	NA	37	0
Editorial board			
Integrative & Comparative Biology	74	25 (34)	4 (5)
Journal of Experimental Biology	52	13 (25)	3 (6)
Zoological Journal of the Linnean Society	20	11 (55)	3 (15)
Journal of Experimental Zoology	67	26 (39)	3 (4)
Frontiers in Ecology and Evolution	74	25 (34)	4 (5)
Mycology	53	15 (28)	8 (15)
Median, %	NA	34	5
Corresponding author			
2018	104	54 (52)	7 (1)
2019	134	59 (44)	16 (12)
2020	119	55 (46)	18 (15)
2021	134	68 (51)	31 (23)
2022	116	66 (57)	13 (11)

Abbreviations: NA, not applicable; PUI, primarily undergraduate institution.

Lead editors of the top 50 journals in 6 disciplines, editorial board members for 6 life science journals, and corresponding authors published in *Integrative and Comparative Biology*.

sketches, and otherwise used text written by authors with a high likelihood of knowing correct pronouns, such as acknowledgments. When collecting data on lead editors, we excluded publications led by large subject-specific editorial teams.

Results This study included 407 lead editors, 340 editorial board members, and 607 corresponding authors. Overall, women accounted for 51% of corresponding authors, 34% of editorial board members, and 37% of lead editors. Between 1979 and 2023, the proportion of women editors increased steadily from a median of 10% in the 1980s to more than 50% in the 2020s. In contrast, the proportion of PUI editors remained largely flat, with a median near 10%, only increasing in the 2020s under the tenure of a PUI lead editor. Inequality was higher for type of institution. Although more US faculty worked at primarily undergraduate institutions (PUIs) than PhD-granting institutions, only 12% of corresponding authors, 5% of editorial board members, and less than 1% of editors in chief worked at PUIs.

Conclusions The observed trends are consistent with our central idea—indirect indicators of inequality that reflect intersectionality, such as type of institution, reveal more severe inequality in academic publishing than direct inequality indicators, such as gender.

References

1. Memon AR, Ahmed I, Ghaffar N, Ahmed K, Sadiq I. Where are female editors from low-income and middle-income countries? a comprehensive assessment of gender, geographical distribution and country's income group of editorial boards of top-ranked rehabilitation and sports science journals. *Br J Sports Med*. 2022;56(8):458-468. doi:0.1136/bjsports-2021-105042

2. Bates DC, Borland E. Fitting in and stalling out: collegiality, mentoring, and role strain among professors in the sciences at a primarily undergraduate institution. *Polymath*. 2014;4(2):50-68.

3. Curtis JW. Faculty gender equity indicators 2021: data report. Accessed July 16, 2025. https://mountainscholar.org/bitstreams/78250ae2-8a5c-4bc0-b43d-4417d53adba9/download

[1]Fresno State University, Fresno, CA, US, umuller@csufresno.edu; [2]Max Planck Institute for Intelligent Systems, Stuttgart, Germany.

Conflicts of Interest Disclosures Ulrike K. Müller is the editor in chief of *Integrative and Comparative Biology*, a journal included in the dataset analyzed for this study. No other disclosures were reported.

Diversity Among Reviewers Assigned to Evaluate a Paper as a Factor in Diversifying Perspective and Improving the Peer Review Process in Computer Science

Navita Goyal,[1] Ivan Stelmakh,[2] Nihar B. Shah,[3] Hal Daumé III[1]

Objective Peer review often involves multiple reviewers to reduce bias and broaden perspectives.[1,2] We examined the role of diversity among reviewers assigned to evaluate a paper as a factor in diversifying perspectives and improving the utility of the peer review process.[3] We proposed 2 desiderata: reviews should cover most contents of the paper (high coverage) and reviews should add information not already present in other reviews (low redundancy). We hypothesized that reviews from diverse reviewers would exhibit higher coverage and lower redundancy.

Design We conducted a cohort study using data from the International Conference on Machine Learning, a top-tier venue in computer science. Data were collected in February to April 2020 and analyzed January to April 2024. Reviewer diversity was defined along 5 binary dimensions: whether reviewers belonged to the same organization, same geographical region, similar seniority levels, similar research topics, or had common coauthors or papers. Organization and geographical data were obtained from reviewer profiles, seniority was based on h-index, coauthorship was determined from Google Scholar publications records, and topics were inferred using a topic model applied to the abstracts of reviewers' prior publications. Seniority and topical diversity

were binarized using median h-index and topical similarity scores, respectively. Coverage was measured by the breadth of review criteria addressed—specifically, the aspects of the paper discussed (eg, summary, motivation, originality, soundness) and the types of arguments used (eg, fact, request, reference, quote)—and lexical and semantic overlap between the reviews and the paper abstract. Redundancy was measured as the lexical and semantic overlap among different reviews. Both measures were computed using natural language processing tools. To isolate the relationship between diversity and review coverage and redundancy, we controlled for potential confounders, including reviewers' expertise (measured as similarity between reviewer's published papers and the submitted paper), reviewer characteristics (organization, location, topic, coauthors, and seniority), and paper content (by comparing diverse and nondiverse slates of reviewers for the same paper). We used a weighted linear regression model with review coverage or redundancy as the dependent variable and diversity and confounding factors as independent variables. Statistical significance was assessed using *t* tests.

Results This study included 4991 submitted papers and 3637 reviewers. Diversity in publication networks and seniority were associated with broader coverage of review criteria (**Table 25-1066**). Topical diversity was associated with a broader coverage of the paper. No significant association was found between coverage and organizational and geographical diversity. Except for geographical diversity, diversity in organizations, seniority, topics, or publications networks were associated with lower redundancy among reviews. Furthermore, publication network-based diversity alone was associated with varying perspectives (that is, low redundancy) within specific review criteria.

Conclusions Our study revealed that, in addition to optimizing individual reviewer expertise, choosing diverse slates of reviewers can lead to better reviews with higher coverage and lower redundancy. The findings are limited to certain aspects of review utility, namely redundancy and coverage, operationalized through specific measures. We recognize that there may be other factors that are crucial, either generally or specific to a particular conference or journal, which we have not addressed.

References

1. Shah NB. Challenges, experiments, and computational solutions in peer review. *Comm ACM*. 2022;65(6):76-87. doi:10.1145/3528086

2. Lee CJ, Sugimoto CR, Zhang G, Cronin B. Bias in peer review. *J Am Soc Inf Sci Technol*. 2013;64(1):2-17. doi: 10.1002/asi.22784

3. Jackson SE, May KE, Whitney K, Guzzo RA, Salas E. Understanding the dynamics of diversity in decision making teams. *Team Effectiveness Decision Making Organ*. 1995;204:261.

[1]University of Maryland, College Park, MD, US, navita@umd.edu; [2]New Economic School, Moscow, Russia; [3]Carnegie Mellon University, Pittsburgh, PA, US.

Conflict of Interest Disclosures Nihar B. Shah 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 research was supported by National Science Foundation award 1942124 and Office of Naval Research award N000142212181.

Role of the Funder/Sponsor The funders played 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 Information We thank the current and former members of the University of Maryland Computational Linguistics and Information Processing laboratory, especially Alexander Hoyle and Connor Baumler, for their useful suggestions and feedback on this work. This study was approved by the University of Maryland Institutional Review Board.

Table 25-1066. Assessment of Reviewer Diversity by Different Dimensions of Review Coverage and Redundancy[a]

Diversity	Papers, No.	Regression coefficient (*P* value)							
		Coverage				Redundancy		Semantic redundancy within review criteria	
		Review criteria		Paper					
		Argument	Aspect	Lexical	Semantic	Lexical	Semantic	Argument	Aspect
Organizational	237	0.004 (.10)	0.004 (.08)	0.007 (.10)	0.004 (.09)	−0.022 (<.001)	−0.008 (.003)	−0.002 (.21)	−0.001 (.33)
Geographical	1432	0.002 (.85)	0 (.76)	0 (.85)	0 (.13)	−0.003 (.69)	−0.001 (.37)	0 (.87)	0 (.16)
Seniority	1127	0.006 (.003)	0.007 (.007)	0 (.95)	−0.002 (0.80)	−0.006 (.006)	−0.002 (.007)	0 (.05)	0 (.03)
Topical	387	−0.131 (.20)	−0.103 (.18)	0.110 (.004)	0.093 (.01)	−0.052 (.01)	−0.029 (.009)	0.007 (.41)	0.009 (.88)
Publication network	314	0.006 (.009)	0.006 (.008)	0.008 (.08)	0.006 (.10)	−0.026 (<.001)	−0.010 (.004)	−0.002 (.004)	−0.001 (.006)

[a]The *P* value threshold was set at .02 obtained after Benjamini-Hochberg multiple testing correction, with a false discovery threshold of .05.

Geographical Representation of Author Country Among Peer Reviewers and Publishing Success at 60 STEM Journals

James M. Zumel Dumlao,[1] Misha Teplitskiy[1]

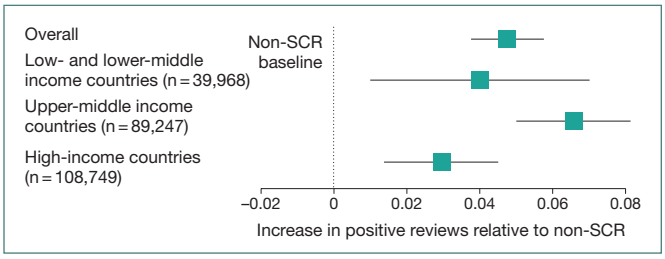
Objective This study tested whether (1) peer reviewers from the same country as the corresponding author are more favorable compared with those from a different country and (2) corresponding authors have differential access to these same-country reviewers (SCRs). While reviewer geographic diversity was previously considered,[1,2] attempts to establish same-country preferences were confounded[2] or had small samples.[3]

Design This cross-sectional study (using the STROBE reporting guideline) took place from October 2022 to May 2025. Using metadata from 204,718 submissions to 60 STEM journals published by the Institute of Physics Publishing from 2018 to 2022, the study employed linear models with fixed effects and *t* tests for significance at the 1% level. Standard errors were clustered at the level of the fixed effects. Only first-round reviews were included to avoid autocorrelation (2.10 first-round reviewers per submission on average; 14.67% were SCRs). Fixed effects limit analysis to within-group comparisons. For example, manuscript fixed effects control for time-invariant variation across manuscripts like quality and proportion of SCRs on the review panel. To test the first hypothesis, this study compared the likelihood of a positive review (outcome) between SCRs and non-SCRs (exposure) using manuscript and reviewer fixed effects. Review positivity is equal to 0 if a reviewer recommended to reject the submission and 1 otherwise. To test the second hypothesis, this study estimated the likelihood a reviewer is an SCR (outcome) by corresponding author country income category (exposure) using journal fixed effects and controls for logged team size and anonymization status. Of 156 corresponding author countries, 60 were high-income countries (HIC), 39 were upper-middle income (UMIC), and 57 were low- and lower-middle income (LLMIC), according to World Bank classification.

Results SCRs were more likely to give positive reviews compared with non-SCRs on the same manuscript (4.78 percentage points higher [95% CI, 3.78-5.78]; n = 181,228). Estimates of relative SCR positivity were positive for all country income groups (**Figure 25-1099**). HIC authors received SCRs over twice as often as LLMIC authors (4.97% [95% CI, 3.50%-6.44%] vs 2.66% [95% CI, 1.41%-3.91%]; n = 237,823). Authors from countries most represented among reviewers—the US, China, and India—had the greatest likelihood of receiving SCRs (30.27% [95% CI, 30.26%-30.28%], 26.59% [95% CI, 24.41%-28.77%], and 13.69% [95% CI, 10.77%-16.61%], respectively; n = 237,823).

Conclusions Lack of geographical diversity among reviewers—possibly reflecting historical patterns of scientific production associated with country wealth—may introduce structural advantages if reviewers favor work from their own country, as this study demonstrated. However, diversification may be challenging if qualified reviewer capacity is low in underrepresented countries. Further research should examine generalizability across fields and test potential remedies.

Figure 25-1099. Same-Country Reviewers (SCRs) Are More Likely to Give Positive Reviews Compared With Non-SCRs on the Same Manuscript



Marginal probability that a review is positive in overall data and country income group subsets. Cluster-robust standard errors were clustered at the level of the fixed effects; error bars indicate 95% CIs. The line at zero represents the positivity likelihood of non-SCRs.

References

1. Smith OM, Davis KL, Pizza RB, et al. Peer review perpetuates barriers for historically excluded groups. *Nat Ecol Evol*. 2023;7(4):512-523. doi:10.1038/s41559-023-01999-w

2. Murray D, Siler K, Larivière V, et al. Author-reviewer homophily in peer review. scientific communication and education *bioRxiv*. 2018. doi:10.1101/400515

3. 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

[1]University of Michigan School of Information, Ann Arbor, MI, US, jamesmzd@umich.edu.

Conflict of Interest Disclosures None reported.

Funding/Support This study received funding from the Institute of Physics Publishing, Schmidt Futures, and Science for Progress Initiative.

Role of the Funder/Sponsor The Institute of Physics Publishing was involved in the collection, management, analysis, and interpretation of the data and preparation, review, or approval of the abstract. Schmidt Futures and Science for Progress Initiative contributed personnel funding.

Acknowledgment We thank the Institute of Physics Publishing, and particularly Kim Eggleton and Mikka Pers, for supporting this project. We thank Schmidt Futures and the Science for Progress Initiative for financial support.

Author Responses to Editorial Guidance on Reporting of Sex, Gender, Race, and Ethnicity Data

Mabel Chew,[1] Taissa Vila,[2] Jashelle Caga-Meller,[3] Zoë Mullan,[4] Diana Samuel[5]

Objective As part of The Lancet Group's commitment to advancing equity, diversity, and inclusion, we implemented the Sex and Gender Equity in Research (SAGER) guidelines[1]

in 2023, and *The Lancet* guidance on reporting race and ethnicity[2] in June 2024 across 24 Lancet Group journals. Although data exist on the reporting of race and ethnicity in published articles after similar guidelines were launched,[3] less is known about how authors view and engage with guidelines, which is crucial to implementing and refining editorial policy. Thus, we aimed to examine authors' awareness of and responses to these guidelines and identify challenges to guideline implementation.

Design Corresponding authors of research or review articles submitted to a *Lancet* journal, who had received a request for revision between July and December 2024, were invited to participate in an online survey in January 2025. The survey asked authors about their awareness of the guidelines, changes they made to their manuscripts in response to editorial advice, the ease with which they made those changes, any challenges encountered, and how these guidelines might influence their next project.

Results The survey response rate was 22% (484 of 2185 invitations). The largest proportion of respondents was from the US (68 [14%]), followed by China (65 [13%]), the UK (59 [12%]), Australia (30 [6%]), the Netherlands (27 [6%]), Germany (21 [4%]), Canada (19 [4%]), and Sweden (17 [4%]). A total of 262 respondents (54%) were men, 312 (64%) were involved in research and/or development, and 191 (40%) were senior researchers or in middle management. Of the 484 respondents, 193 (40%) and 220 (46%) were aware of the sex and gender and race and ethnicity reporting guidelines before submitting, respectively. A total of 246 respondents (51%) and 266 (55%), respectively, were not required to collect these data, and 9 (2%) and 33 (7%) were not permitted to collect these data. Among 153 respondents (32%) who amended their manuscript as a result of the sex and gender guidance, 104 found this easy or very easy to do. Fewer respondents (90 [19%]) made changes in response to the race and ethnicity guidance, with 55 finding this easy or very easy. **Box 25-1122** includes open-text responses on challenges in addressing these guidelines. Approximately one-half of respondents said they were likely or very likely to do things differently in their next project as a result of these guidelines (255 [53%] and 221 [46%] for each guidance, respectively).

Conclusions Less than one-half of the authors surveyed reported awareness of sex and gender or race and ethnicity reporting guidelines before submission, suggesting that increasing awareness could enhance engagement. Most authors who amended their manuscript in response to these guidelines found this to be easy. The challenges identified provide opportunities for journals to refine editorial processes.

Box 25-1122. Challenges in Addressing Reporting Guidance (Open Responses)

Sex and gender

Common reasons for not making changes related to this guidance

- The manuscript already adhered to guidelines
- Guidelines were not applicable to their data (eg, data did not involve humans or participants were all of 1 sex
- Editors or reviewers did not ask for these changes

Common challenges when making changes related to this guidance

- Limitations of the data itself (eg, incomplete or no sex or gender data in the original dataset
- Extra time and effort required for further data extraction or analysis
- Word limits
- Geopolitical events
- A total of 22 of 58 respondents stated that there were no challenges

Race and ethnicity

Common reasons for not making changes related to this guidance

- Guidelines were not applicable to their data (eg, no human participants or a nondiverse population
- Race or ethnicity data were not available, collected, or analyzed
- The manuscript already adhered to guidance
- Editors or reviewers did not ask for these changes

Common challenges when making changes related to this guidance

- Limitations of the data (eg, unclear, inconsistent, or outdated definitions or categorization for race or ethnicity)
- Extra time and effort required
- A total of 12 of 27 respondents stated that there were no or minimal challenges

References

1. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. *Res Integr Peer Rev*. 2016;1:2. doi:10.1186/s41073-016-0007-6

2. Chew M, Samuel D, Mullan Z, Kleinert S; Lancet Group for Racial Equity (GRACE). The Lancet Group's new guidance to authors on reporting race and ethnicity. *Lancet*. 2024;403(10442):2360-2361. doi:10.1016/S0140-6736(24)01081-X

3. Flanagin A, Cintron MY, Christiansen SL, et al. Comparison of reporting race and ethnicity in medical journals before and after implementation of reporting guidance, 2019-2022. *JAMA Netw Open*. 2023;6(3):e231706. doi:10.1001/jamanetworkopen.2023.1706

[1]*The Lancet,* Elsevier Australia, Chatswood, Sydney, Australia, mabel.chew@lancet.com; [2]*The Lancet Regional Health—Americas,* Rio de Janeiro, Brazil; [3]*The Lancet Regional Health—Western Pacific*, Sydney, Australia; [4]*The Lancet Global Health* London, UK; [5]*The Lancet Digital Health*, London, UK.

Conflict of Interest Disclosures Mabel Chew is a member and former co-chair of The Lancet Group for Racial Equity and has

received funding for conference travel expenses from the World Conference on Research Integrity, Committee on Publication Ethics, and Nuffield Department of Primary Care Health Sciences. Taissa Vila is co-chair of and a member of The Lancet Group for Racial Equity. Jashelle Caga-Meller is honorary clinical senior lecturer at the Faculty of Medicine and Health, University of Sydney, Australia, and a member of The Lancet Group for Racial Equity. Diana Samuel is a member and former co-chair of The Lancet Group for Racial Equity, and a member and former Chair of the European Association of Science Editors' (EASE) EDI Committee.

Acknowledgments The authors wish to thank Louise Hall and Adrian Mulligan from the Elsevier Customer Insights team for their help in developing and conducting the survey and for their support in data reporting and analysis and *The Lancet* colleagues Pooja Jha, Rupa Sarkar, Lan-Lan Smith, and Richard Horton for their advice and support of this project.

Representation of Authors From Low- and Middle-Income Countries (LMICs) in Trials With Participants From LMICs

Harleen K. Marwah,[1] Abarna Pearl,[1] Christos P. Kotanidis,[1,2,3] Sarah Gorey,[1,4] Darren Taichman,[1,5] Mary Beth Hamel[1,6]

Objective Trials increasingly include participants from low- and middle-income countries (LMICs),[1] but authorship representation may not be proportional. LMIC-affiliated authors on trials recruiting from LMICs offer important perspectives, which have implications for equity, expertise, and capacity building. Prior data have looked at specialty journals or non-Western country categories,[2] but author representation specifically in trials with participants from LMICs is not known. This study evaluated trends over time in LMIC-affiliated authorship among trials recruiting participants from LMICs.

Design Using PubMed, we extracted trials published during alternating months (6 mo/y) in the years 2008, 2013, 2018, and 2023 from 4 high-impact medical journals: the *New England Journal of Medicine*, *The Lancet*, *JAMA*, and *BMJ*. Four investigators manually identified trials that recruited participants from LMICs, defined by the Wellcome Trust as countries with low-income or middle-income economies.[3] We designed and calculated 3 metrics for each trial: (1) an enrolling country ratio (the number of recruiting LMICs divided by the total number of recruiting countries), (2) an author ratio (the number of LMIC-affiliated authors divided by the total number of authors in a trial), and (3) a representation index (RI) (the author ratio divided by the enrolling country ratio). The RI controlled for the proportion of recruiting LMICs in each trial. An RI less than 1 indicated that the proportion of LMIC-affiliated authors was lower than the proportion of recruiting LMICs. An RI of 1 indicated that the 2 ratios were exactly proportionate.

Results The PubMed search yielded 865 trials, 302 of which recruited participants from LMICs. Of these, 139 (46.0%) were published in the *New England Journal of Medicine*, 117 (38.7%) in *The Lancet*, 35 (11.6%) in *JAMA*, and 11 (3.6%) in *BMJ*. The number and proportion of trials recruiting participants from LMICs increased between 2008 (52/174, 29.9%) and 2023 (117/262, 44.7%). The mean enrolling country ratio decreased over time, from 0.62 in 2008 to 0.41 in 2023 (**Figure 25-1131**, blue line). The mean author ratio decreased after 2008 and then remained similar; in 2008, it was 0.34 and in 2023, it was 0.23 (**Figure 25-1131**, gray line). Between 2008 and 2023, the RI increased modestly from 0.49 to 0.53 (**Figure 25-1131**, teal line).

Conclusions This study highlights that LMIC-affiliated authorship in LMIC-recruiting trials has increased modestly over time but remains disproportionate to LMIC participant recruitment sites. Limitations of the study include the lack of information on the proportion of LMIC-recruited participants for each trial. This information could be a stronger indicator as to what LMIC-affiliated authorship representation should look like. Also, LMIC definitions may change over time and by defining organization. While there has been modest improvement in author representation, increased attention may help promote representative and ethical research.

References

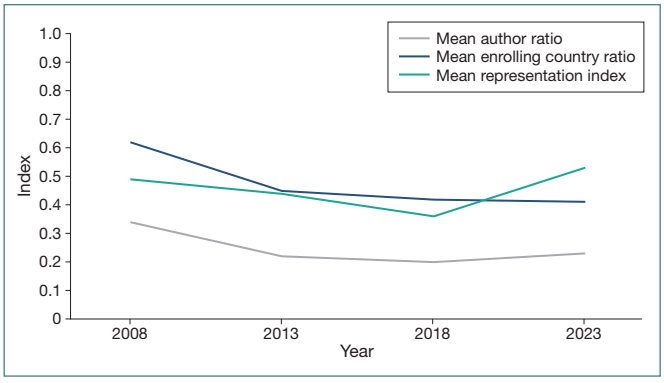
1. Number of clinical trials by year, country, WHO region and income group (1999-2024). World Health Organization. December 2024. Accessed April 15, 2025. https://www.who.int/observatories/global-observatory-on-health-research-and-development/monitoring/number-of-clinical-trials-by-year-country-who-region-and-income-group

2. De Oliveira-Gomes D, Guilliod C, Gebran K, et al. Equity and representation in cardiology research: a comprehensive analysis of authorship from low and low-middle income countries in cardiology journals. *Curr Probl Cardiol*. 2024;49(8):102667. doi:10.1016/j.cpcardiol.2024.102667

3. Low- and middle-income countries. Wellcome Trust. Accessed January 8, 2025. https://wellcome.org/grant-funding/guidance/prepare-to-apply/low-and-middle-income-countries

[1]*New England Journal of Medicine*, Boston, MA, US, hmarwah@nejm.org; [2]Heart and Vascular Center, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US; [3]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, UK; [4]Stroke Clinical Trials Network Ireland, University College Dublin Clinical Research Centre, Mater Misericordiae University Hospital, Dublin, Ireland; [5]Department of Medicine, Division of Pulmonary, Allergy and Critical Care Medicine, University of Pennsylvania School of

Figure 25-1131. Mean Calculated Indices per Trial, by Year



The mean representation index was calculated from the average representation

Medicine, Philadelphia, PA, US; [6]Division of General Medicine and Primary Care, Department of Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA, US.

Conflict of Interest Disclosures Harleen K. Marwah, Abarna Pearl, Christos P. Kotanidis, and Sarah Gorey are editorial fellows at the *New England Journal of Medicine* (2024-2025). Darren Taichman is a deputy editor and Mary Beth Hamel is executive editor of the *New England Journal of Medicine*. No other disclosures were reported.

Evolution of Authorship Diversity in African Surgical Research Over 2 Decades

Vincent Kipkorir,[1] Godfrey Sama,[1] Mumba Chalwe,[1] Tihitena Negussie,[1] Michael Mwachiro,[1] Robert Parker,[1] Seke Kazuma,[1] Stella Itungu,[1] Abebe Bekele[1]

Objective Underrepresentation along geographical and gender lines has historically been observed in academic surgery.[1] However, efforts have been made in recent years to ensure equity and inclusion in the surgical learning environment to increase female representation and ensure geographically diversified research engagements.[2] We set out to explore the evolution of authorship diversity in African surgical research using publications featured in an African surgical journal.

Design This was a retrospective cross-sectional study. Data were collected from all articles published in the *East and Central African Journal of Surgery* between 2003 and 2023. Article metadata and author biodata were extracted from each article for subsequent analysis. Author gender was established by 2 independent reviewers through an online search, confirmed by a minimum of 2 web pages including LinkedIn, editorial profiles, institutional websites, and ResearchGate. For original articles, case series, and case reports, geographical location was determined by the primary study site. For the other articles, we used the author's institutional affiliation.

Results A total of 889 articles were included featuring 3001 authors; 2609 (86.9%) were male and 392 (13.1%) were female. In total, 92 females (3.1%) were designated as first authors, 93 (3.1%) as corresponding authors, and 90 (3.0%) as last authors. There was a predominance of males in the positions of first, last, and corresponding author. Despite this, the number of females represented in these categories had a steady increase over the years. The gender disparity ratio increased in the first 15 years, then dropped in the last 5 years. There were 2523 African authors (84.1%), with Ethiopia having the largest number of authors (155 [5.2%]). The distribution according to country and the trends in female authorship are presented in **Figure 25-1149**.

Conclusions Our findings show disparities in the ratio of male to female authors, with a substantial male predominance. While it is encouraging to note the steady increase in the number of female authors over the years, a wide gap remains. The steady rise may be attributed to the overall rise in female surgeons following the recent establishment of women in surgery associations and better advocacy and funding toward empowerment of female surgeons in Africa. With a female surgical workforce density of about 11% in College of Surgeons of East, Central and Southern Africa member countries, further efforts are needed to ensure that these potential female authors are well represented, especially as first, last, and corresponding authors.[3] Strategies currently in place to foster surgical research in Africa need to be sustained to maintain the current trajectory of African representation in the global surgical research landscape. This study is limited by the subjective nature of ascertaining gender and location; however, the level of bias was reduced by having 2 independent assessors.

References

1. Rathna RB, Biswas J, D'Souza C, Joseph JM, Kipkorir V, Dhali A. Authorship diversity in general surgery-related Cochrane systematic reviews: a bibliometric study. *BJS*. 2023;110(8):989-990. doi:10.1093/bjs/znad117

2. Burks CA, Russell TI, Goss D, et al. Strategies to increase racial and ethnic diversity in the surgical workforce: a state of the art review. *Otolaryngol Head Neck Surg*. 2022;166(6):1182-1191.

3. Osei-Kuffour D, Banda CH, Campion A, et al. The evolution of the specialist surgeon workforce in East, Central and Southern Africa. World J Surg. 2025;49(4):946-954. doi:10.1002/wjs.12545

[1]*East and Central African Journal of Surgery*, College of Surgeons of East, Central and Southern Africa; managing_editor@cosecsa.org.

Conflict of Interest Disclosures None reported.

Virtual

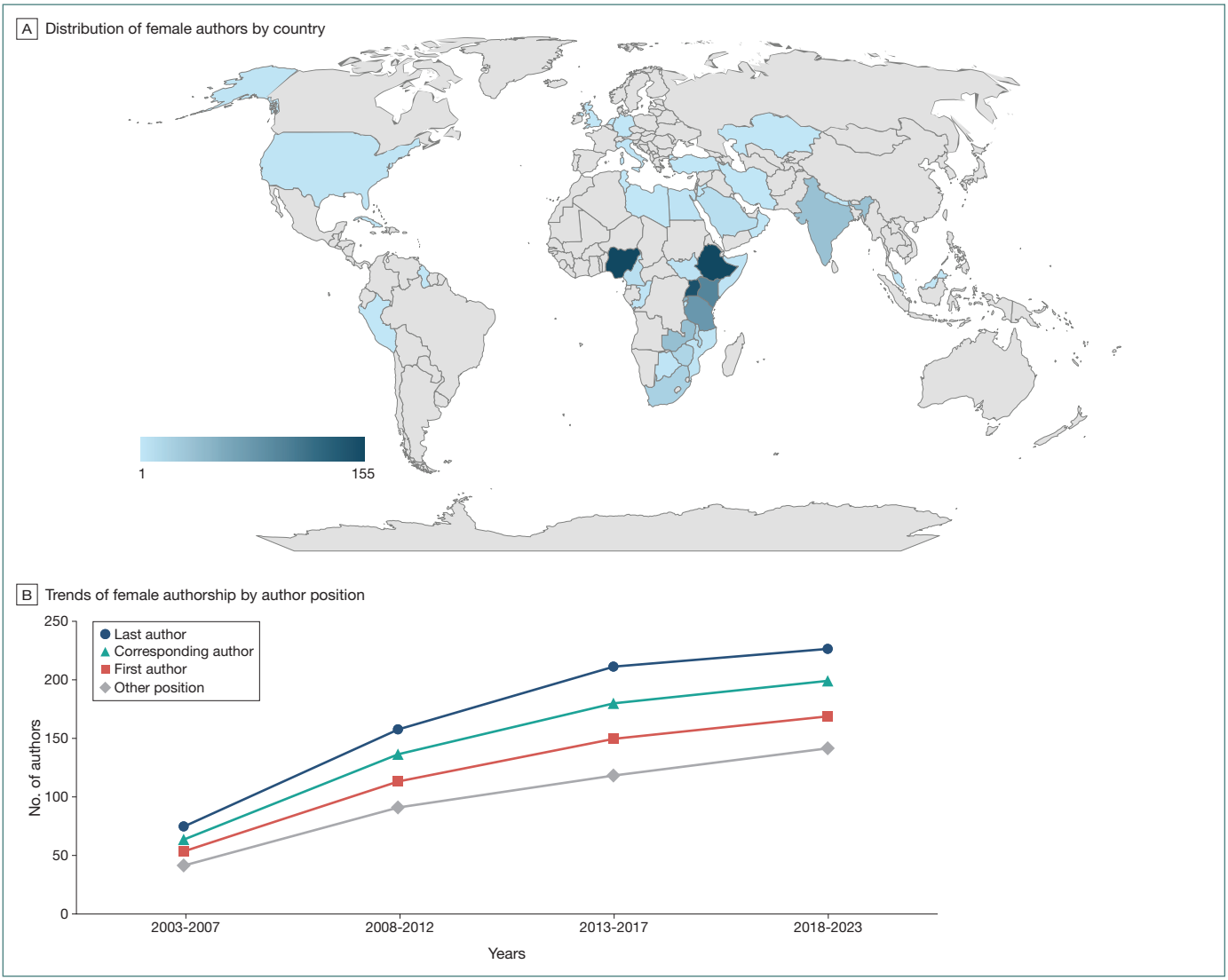
Integrating Indigenous Knowledge Into Peer Review Processes in Nigerian Environmental and Health Research

Oludele M. Solaja[1]

Objective Scientific peer review often excludes Indigenous knowledge (IK), favoring Western scientific norms. In Nigeria, valuable local insights on health and environmental practices are marginalized, limiting their impact and perpetuating epistemic injustice.[1] Integrating IK into peer review can enhance research relevance, cultural acceptance, and credibility, providing region-specific solutions crucial for sustainable development in Nigerian communities. This research examines how IK can be incorporated into scientific peer review to remove epistemological racism and improve the representation and practicality of IK in Nigerian environment and health research.[2]

Design The study follows a 3-phase approach: (1) literature search, (2) stakeholder consultation, and (3) development and pilot testing of guidelines. In phase 1, a comprehensive literature review of 142 peer-reviewed articles (identified via PubMed, Scopus, and African Journals Online) was

Figure 25-1149. Female Authorship in the *East and Central African Journal of Surgery,* 2003-2023



conducted to identify gaps in IK recognition within environmental and health research in Nigeria. Phase 2 included 396 surveys, 30 semistructured interviews, and 3 focus group discussions involving Nigerian researchers, peer reviewers, and IK holders. This engagement was conducted to highlight cultural barriers and biases in existing peer review practices. Phase 3 involved developing standardized guidelines based on stakeholder feedback to incorporate IK into peer review, followed by pilot testing with a subset of 12 articles from 3 Nigerian journals (*Nigerian Journal of Environmental Sciences and Technology*, *FUTA Journal of Research in Sciences*, *Ibadan Journal of the Social Sciences*) from April to June 2025 to assess feasibility and effectiveness. In this phase, structured reviewer and editorial feedback was collected and is currently being analyzed to evaluate key metrics, including reviewer satisfaction (measured through structured postreview surveys), inclusivity ratings (assessed via 5-point Likert scale responses and demographic tracking of reviewer diversity), and manuscript quality (evaluated by independent editorial panels comparing clarity, cultural sensitivity, and scientific rigor between IK-inclusive and conventional reviews).

Results Among the 142 articles reviewed, only 18 (12.7%) referenced IK, and none demonstrated structured inclusion of IK in peer review or evaluation criteria. In total, 32 of 38 IK stakeholders (84.2%) and 50 of 64 researchers (78.1%) perceived current peer review practices as exclusionary, citing language bias, absence of culturally grounded criteria, and a lack of reviewers who were knowledgeable in IK.[3] Pilot testing yielded the following outcomes: 19 of 24 reviewers (79.2%) reported increased clarity and confidence in evaluating IK-informed manuscripts; inclusivity ratings increased from a baseline mean of 2.8 to 4.3 on the 5-point Likert scale; and independent editorial panels reported that IK-inclusive reviews scored 25% to 30% higher in cultural sensitivity and contextual relevance without compromising scientific rigor. Stakeholders reported that integrating IK leads to a more holistic and locally resonant understanding of health and environmental challenges in Nigeria.

Conclusions Integrating IK into peer review can make research validation more equitable and locally relevant, aligning with global calls for inclusivity in scientific processes. This framework, grounded in empirical findings, may serve as a model for adapting peer review standards to respect diverse

knowledge systems and improve the societal impact of research.

References

1. Agrawal A. Dismantling the divide between indigenous and scientific knowledge. *Dev Change*. 1995;26(3):413-439. doi:10.1111/j.1467-7660.1995.tb00560.x

2. Smith LT. *Decolonizing Methodologies: Research and Indigenous Peoples*. Zed Books; 1999.

3. Nakashima D, Roué M. Indigenous knowledge, peoples, and sustainable practice. In: Timmerman P, ed. *Encyclopedia of Global Environmental Change*. John Wiley & Sons; 2002:314-324.

[1]Department of Sociology, Olabisi Onabanjo University, Ago-Iwoye, Nigeria, solaja.oludele@oouagoiwoye.edu.ng.

Conflict of Interest Disclosures None reported.

Geographical Disparities in Navigating Rejection in Scientific Publications

Hong Chen,[1] Christopher I. Rider,[2] David Jurgens,[1,3] Misha Teplitskiy[1]

Objective This study examines which intended scientific contributions become published contributions. While many manuscripts are rejected, there is limited understanding of which researchers successfully navigate rejection and resubmission. We investigate geographical disparities in postrejection publishing outcomes, focusing on differences between authors from Western and non-Western countries. We investigate several potential mechanisms underlying these disparities, especially procedural knowledge of how to interpret editorial decisions for rejection, revision, and resubmission.

Design Partnering with the Institute of Physics Publishing, we collected metadata on 126,000 rejected submissions made in 2018 through 2022 from 62 peer-reviewed journals. We tracked whether, when, and where these rejected manuscripts were eventually published by developing a classifier that we validated with ground-truth data provided by the authors of rejected manuscripts. We defined Western authors as those affiliated with (self-reported) institutions in North America, Europe, or Oceania, and non-Western authors as those from Asia, Africa, and Latin America. The analysis examined the publishing disparities based on author geography while controlling for key confounders—perceived submission quality, proxied by peer reviewer evaluations at the time of rejection (including desk rejection)—as well as time, team composition, and corresponding author characteristics, such as gender, seniority, and current institutional affiliation ranking. The prior Western coauthors and Western affiliations were also included, serving as proxies for access to procedural knowledge. To further probe procedural knowledge as a mechanism, we examined multiple dimensions of how authors interpret and respond to rejection in a survey of corresponding authors whose manuscripts were rejected.

Results Among the authors included, 43% had Western affiliations and 57% had non-Western affiliations. Authors from Western countries had better postrejection outcomes. Despite similar perceived quality at the time of rejection, Western authors were 6.7% more likely to publish the rejected papers elsewhere. Among those who ultimately published (59.3% of all rejected papers), Western authors published approximately 23 days faster and did 5.9% less abstract revising and 12.0% less authorship team change. We also found evidence suggesting that greater access to procedural knowledge, proxied by prior publishing experience and prior Western affiliation and coauthorship, contributed to better postrejection outcomes. These factors explain a substantial portion of the observed disparities in postrejection publication rates, more so than perceived submission quality or institutional resourcefulness. We also surveyed 10,000 corresponding authors of rejected manuscripts in June 2023 and received 287 responses. Limited by the very low response rate, survey responses indicated no significant differences between Western and non-Western authors in their perceptions of feedback fairness or negativity, planned revisions, or target outlet prestige, suggesting that other dimensions of procedural knowledge might be key.

Conclusions This study identified geographical disparities in the file drawer of science. Western authors experienced better postrejection outcomes than non-Western authors, and the findings suggest that differential access to procedural knowledge may contribute to these disparities.

[1]School of Information, University of Michigan, Ann Arbor, MI, US, tepl@umich.edu; [2]Ross School of Business, University of Michigan, Ann Arbor, MI, US; [3]Department of Computer Science and Engineering, University of Michigan, Ann Arbor, MI, US.

Conflict of Interest Disclosures None reported.

Editorial Landscape of Journals in Kenya, Ethiopia, Nigeria, and Mozambique

Patrick Amboka,[1] Tony Blair Aloo,[1] Daniel Krugman,[1,2] Abel Simiyu,[1] Hiram Kariuki,[1] Benard Ondiek,[1] Nosa Orobaton,[3] Emmy Igonya,[1] Alphonsus Neba,[1] Marta Vicente-Crespo,[1,4] Julius Kirimi Sindi[1]

Objective This study explored editorial practices among African journals, examined the factors influencing these practices, and assessed authors' perspectives and preferences in selecting journals for publication. A large proportion of African journals (63.2%) are neither discoverable via Google Scholar nor included in Scopus.[1] However, their discoverability and inclusion in international platforms are nuanced and not always straightforward; hence, quality markers need to be better aligned.

Design This study triangulated multiple sources of information and used qualitative data gathering techniques to capture nuances and provide deeper insights into the performance and visibility of African journals. Conducted between July and December 2024, the study utilized a cross-sectional design. In-depth interviews (n = 60), key

informant interviews (n = 32), and focus group discussions (n = 7) were conducted in Kenya, Ethiopia, Nigeria, and Mozambique. A purposive sampling technique was used to identify participants. Ethical approvals were obtained from relevant institutions. We obtained consent from participants to record and publish the data in anonymized form. Qualitative data from the audio-recorded interviews were transcribed using Microsoft Word and exported to NVivo software for analysis. The data were analyzed using a reflexive thematic analysis.[2] Emerging concepts from the qualitative data were also analyzed using Leximancer software.

Results Four main themes emerged from participant interviews in this study. First, there are challenges in adhering to international editorial standards ("We often want to follow international best practices, but there's limited guidance and mentorship available to help us meet those expectations consistently"). Second, financial constraints undermine editorial integrity ("Sometimes, we're forced to accept submissions not based on quality, but because we need the processing fees to keep the journal running. It's not ideal, but without funding, we have few options"). Third, technological and digital infrastructure gaps exist ("Our journal management system is outdated, and we lack IT support to upgrade it"). Fourth, authors have perceptions of inferiority and trust issues ("Authors often tell us they prefer to submit to Western journals because they assume ours won't be taken seriously").

Conclusions Addressing financial constraints, peer review inefficiencies, and infrastructure gaps is critical to strengthening the editorial practices and credibility of African journals. Overcoming historical neglect through sustainable funding, capacity building, and technological advancement will be key to enhancing their global visibility, trust, and academic impact.

References

1. Amboka P, Sindi JK, Wamukoya M, et al. Discoverability of African journals by Google Scholar and inclusion in Scopus. *VeriXiv*. Preprint posted online November 1, 2024. https://verixiv.org/articles/1-17

2. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Qual Quant*. 2022;56:1391-1412. doi:10.1007/s11135-021-01182-y

[1]African Population and Health Research Center, Nairobi, Nairobi County, Kenya, pamboka@aphrc.org; [2]Department of Anthropology, Brown University, Providence, RI, US; [3]Bill & Melinda Gates Foundation, Seattle, WA, US; [4]School of Public Health, University of the Witwatersrand, Johannesburg, South Africa.

Conflict of Interest Disclosures None reported.

Funding/Support This study was funded by the Bill and Melinda Gates Foundation (grant INV-034355).

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

Editorial and Peer Review Processes

In-person

Effects of Peer Review and Editorial Workflows in Decision-Making at a Diamond Open Access Journal

Tais Freire Galvão,[1,2] Everton Nunes da Silva,[2] Jorge Otávio Maia Barreto,[3] Marcus Tolentino Silva[4]

Objective Detailed descriptions of journals' workflows exist,[1] but empirical studies assessing their effects are scarce. In 2024, a new workflow was implemented at *Epidemiologia e Serviços de Saúde: revista do SUS* (*RESS*) aimed at getting timely and quality reviews. This study assessed the effect of a new workflow on the journal's time to decision-making.

Design This before and after study using historical controls assessed submissions to *RESS* from June to December 2024 (intervention) and June to December 2023 (control), followed up until January 2025. The intervention started in June 2024, based on inclusion of a peer review administrator in charge of reviewers' assignments and adding new and trained associate editors[2]; previous evaluation by the editor in chief (scope) and scientific editor (methodological quality); and checklists to standardize reviews.[3] The comparator was the previous workflow, centralized on the associate editor's assignment of reviewers, and final decisions made by the editor in chief and a member of the editorial committee. Outcomes were processing time from submission to decision, number of reviewers invited, and reviews received. We summarized means and SDs or medians and IQRs. Normality was tested using the Shapiro-Wilk test and square root or logarithmic transformations preceded t-tests, with the Levene test used to assess variances' homogeneity. In unsuitable t-tests, the Mann-Whitney test was applied. Residual normality and heteroscedasticity were examined by Breusch-Pagan test. We adjusted by the length of the paper (full or short) to minimize potential confounding.

Results Out of 928 submissions in 2024, 218 proceeded to peer review (140 ongoing until cohort closure) and 78 had a decision (36 accepted, 42 rejected). In 2023, from 970 submissions, 77 were peer reviewed and resulted in 28 acceptances and 49 rejections ($P = .22$). Mean (SD) time in days to decision was significantly shorter in the intervention (62.8 [40.5]) than in the controls (130.1 [64.5]; $P < .001$), which was also faster in the intervention according to acceptance (92.8 [39.8] vs 133.5 [61.6]; $P = .002$) or rejection (37.1 [16.0] vs 128.1 [66.7]; $P < .001$). The mean (SD) number of invited reviewers per manuscript was higher in the intervention (10.1 [6.3]) than in the controls (7.8 [5.5]; $P = .02$), without significant difference according to decision when compared with the historical controls (approved, 10.0 [7.0] vs 6.3 [2.8]; $P = .07$; rejected, 10.3 [5.8] vs 8.6 [6.5]; adjusted $P = .11$). The median (IQR) number of reviews received was also higher (4 [4-4] vs 2 [2-3]; $P < .001$), and was similarly higher when comparing decisions across groups

(approved, 4 [4-4] vs 2 [2-3]; $P < .001$; rejected, 4 [4-4] vs 2 [2-3]; $P < .001$) (**Table 25-0909**).

Conclusions The new workflow reduced the duration of editorial processing and allowed a higher number of peer reviews to inform decision-making. The outcomes were not defined before the intervention was planned; the length of follow-up, new editors, and submission system were confounders that may have influenced the findings.

References

1. Mark H, Ragon T, Funning G, et al. Editorial workflow of a community-led, all-volunteer scientific journal: lessons from the launch of Seismica. *Seismica*. 2023;2(2):1091. doi:10.26443/seismica.v2i2.1091

2. Galvão TF, Silva EN, Araújo WN, Barreto JOM. Editorial improvement in *Epidemiologia e Serviços de Saúde* in 2024. *Epidemiol Serv Saude*. 2024;33:e20241002. doi:10.1590/S2237-96222024v33e20241002.en

3. Silva MT, Galvão TF. Systematization of peer review in *Epidemiologia e Serviços de Saúde*. *Epidemiol Serv Saude*. 2024;33:e20241001. doi:10.1590/S2237-96222024v33e20241001.en

[1]Faculdade de Ciências Farmacêuticas, Universidade Estadual de Campinas, Campinas, Brazil, taisgalvao@gmail.com; [2]Secretaria de Vigilância em Saúde e Ambiente, Ministério da Saúde, Brasília, Brazil; [3]Faculdade de Ciências e Tecnologias em Saúde, Universidade de Brasília, Brasília, Brazil; [4]Gerência Regional de Brasília, Fundação Oswaldo Cruz, Brasília, Brazil; [5]Faculdade de Ciências da Saúde, Universidade de Brasília, Brasília, Brazil.

Table 25-0909. Outcomes and Time to Decision According to the New or Former Workflow of *Epidemiologia e Serviços de Saúde: revista do SUS*

	New workflow	Former workflow	*P* value	Adjusted *P* value[a]
Manuscripts included, No.	78	77		
Accepted	36	28	.22[b]	
Rejected	42	49		
Length of manuscript				
Full (~3500 words), No.	66	71	.30[b]	
Short (~1500 words), No.	11	7		
Days to decision, mean (SD)	62.8 (40.5)	130.1 (64.5)	<.001[c]	<.001
Accepted	92.8 (39.8)	133.5 (61.6)	.002[d]	.002
Rejected	37.1 (16.0)	128.1 (66.7)	<.001[e]	<.001
Peer reviewers invited, mean (SD)	10.1 (6.3)	7.8 (5.5)	.02[e]	.02
Accepted	10.0 (7.0)	6.3 (2.8)	.07[f]	.07
Rejected	10.3 (5.8)	8.6 (6.5)	.13[e]	.11
Peer reviews received, median (IQR)	4 (4-4)	2 (2-3)	<.001[g]	<.001
Accepted	4 (4-4)	2 (2-3)	<.001[g]	<.001
Rejected	4 (4-4)	2 (2-3)	<.001[g]	<.001

[a]Adjusted by length of manuscripts, using linear regression.
[b]Calculated with Pearson χ^2 test.
[c]Calculated with t-test with equal variances of square root.
[d]Calculated with t-test with equal variances.
[e]Calculated with t-test with equal variances of logarithm.
[f]Calculated with t-test with equal variances of inverse of square root.
[g]Calculated with Mann-Whitney test.

Conflict of Interest Disclosures Tais Freire Galvão, Everton Nunes da Silva, and Jorge Otávio Maia Barreto are editors of *Epidemiologia e Serviços de Saúde: revista do SUS*. Marcus Tolentino Silva has no conflicts of interest to declare. Tais Freire Galvão is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Reminding Peer Reviewers to Comment on Reporting Items as Instructed by the Journal: An Analysis of 2 Randomized Trials

Hillary Wnfried Ramirez,[1,2,3] Malena Chiaborelli,[1,2,3] Christof M. Schönenberger,[1] Katie Mellor,[4,5] Alexandra N. Griessbach,[1] Paula Dhiman,[4,6] Pooja Gandhi,[7] Szimonetta Lohner,[8,9] Arnav Agarwal,[10,11] Ayodele Odutayo,[5,12] Michael M. Schlussel,[4,6] Philippe Ravaud,[13,14] David Moher,[15,16] Matthias Briel,[1,10] Isabelle Boutron,[13,14] Sally Hopewell,[4] Sara Schroter,[17,18] Benjamin Speich[1,4]

Objective Two randomized controlled trials (RCTs) conducted at journal level have shown that reminding peer reviewers about the 10 most important and underreported reporting items did not improve the reporting quality in published articles.[1] With this pooled in-depth analysis of peer reviewer reports, we aimed to assess at what stage the intervention failed.

Design A subsample of peer reviewer reports from the control group (receiving no reminder) and the intervention group (receiving a reminder of the 10 most important reporting items) were analyzed. In brief, 2 blinded authors independently extracted from peer reviewer reports how many of the 10 key reporting items were flagged by peer reviewers for clarification. The main outcome of this analysis was the mean proportion of the 10 selected reporting items for which at least 1 peer reviewer requested clarification, assessed at the manuscript level. Furthermore, we assessed how many requested changes were later adequately reported in published articles.

Results Across the RCTs, we had access to peer reviewer reports for 533 manuscripts (265 in the intervention group, assessing comments from 740 peer reviewers; and 268 in the control group, assessing comments from 719 peer reviewers). Our results indicate that reviewers in the intervention group requested clarification on more reporting items than those in the control group. Overall, reviewers in the intervention group flagged 21.1% of the 10 reporting items for clarification compared with 13.1% in the control group (mean difference, 8.0 percentage points (pp); 95% CI, 4.9-11.1 pp). However, the overall mean difference between groups was diluted from 8.0 to 4.2 pp when only assessing accepted and published articles and decreased even further to 2.6 pp when only considering changes that were then implemented by authors of manuscripts (**Table 25-0956**). Approximately 55% of reporting items that were criticized by peer reviewers were

Table 25-0956. Comparison of Proportions of Reporting Items That Peer Reviewers Asked to Have Clarified and Which Were Later Adequately Reported in the Published Articles for CONSORT-PR and SPIRIT-PR Combined

Outcome	Requested to be clarified					
	Overall		Accepted and published manuscripts		Adequately reported in published articles	
	Intervention group (n = 265)	Control group (n = 268)	Intervention group (n = 150)	Control group (n = 152)	Intervention group (n = 150)	Control group (n = 162)
Proportion of items for which clarification was requested, No./total No. (%)[a]	560/2650 (21.1)	352/2680 (13.1)	308/2650 (11.6)	197/2680 (7.4)	173/2650 (6.5)	105/2680 (3.9)
Difference in proportions between groups, percentage points	8.0		4.2		2.6	
Reporting items correctly reported in article when requested by peer-reviewer, No./total No. (%)[b]	NA	NA	NA	NA	173/308 (56.2)	105/197 (53.3)

Abbreviation: NA, not applicable.

[a]Referring to the 10 most important and underreported reporting items that were assessed in submitted manuscripts.

[b]The denominator only includes accepted and published manuscripts for which a clarification was requested.

later adequately reported in the published article (intervention group, 173 of 308 [56.2%]; control group, 105 of 197 [53.3%]).

Conclusions Reminding peer reviewers to check reporting items increased their focus on reporting guidelines, leading to more reporting-related requests in their review reports. However, the effect was diluted during the peer review process (particularly due to rejected articles and requests not being implemented by authors). Journals should therefore make sure that requested clarifications are adequately addressed in revised manuscripts.

Reference

1. Speich B, Mann E, Schönenberger CM, et al. Reminding peer reviewers of reporting guideline items to improve completeness in published articles: primary results of 2 randomized trials. *JAMA Netw Open*. 2023;6(6):e2317651.

[1]CLEAR Methods Center, Division of Clinical Epidemiology, Department Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland, benjamin.speich@usb.ch; [2]Swiss Tropical and Public Health Institute, Basel, Switzerland; [3]University of Basel, Basel, Switzerland; [4]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK; [5]Clinical Outcomes Assessment, Clarivate, London, UK; [6]The EQUATOR Network, Oxford, UK; [7]Department of Communication Sciences and Disorders, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AL, Canada; [8]Cochrane Hungary, Medical School, University of Pécs, Pécs, Hungary; [9]MTA–PTE Lendület "Momentum" Evidence in Medicine Research Group, Department of Public Health Medicine, Medical School, University of Pécs, Pécs, Hungary; [10]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada; [11]Division of General Internal Medicine, Department of Medicine, McMaster University, Hamilton, ON, Canada; [12]Division of Nephrology, Toronto General Hospital, University Health Network, Toronto, ON, Canada; [13]Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique Hôpitaux de Paris, Paris, France;[14]Université de Paris, CRESS, Inserm, INRA, Paris, France; [15]Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada; [16]Faculty of Medicine, School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada; [17]*The BMJ*, London, UK; [18]Faculty of Public Health & Policy, London School of Hygiene & Tropical Medicine, London, UK.

Conflict of Interest Disclosures Benjamin Speich and Matthias Briel reported receiving unrestricted grants from Moderna for studies unrelated to the presented work. Sara Schroter is employed by BMJ Publishing Group. Katie Mellor is employed by Clarivate. David Moher, Sally Hopewell, and Isabelle Boutron are members of the Consolidated Standards for Reporting Trials (CONSORT) executive board and authors of the CONSORT 2010 statement. David Moher is an author of the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) 2013 statement. David Moher, Michael M. Schlussel, Paula Dhiman, and Philippe Ravaud are members of the Enhancing the Quality and Transparency of Research (EQUATOR) network. Isabelle Boutron and David Moher are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract. No other disclosures were reported.

Funding/Support Benjamin Speich was supported by a Return Postdoc.Mobility (P4P4PM_194496) grant from the Swiss National Science Foundation. Christof M. Schönenberger was funded by the Janggen Pöhn Foundation and the Swiss National Science Foundation (MD-PhD grant No. 323530_221860). Szimonetta Lohner was supported by the Hungarian Academy of Sciences (MTA) within the framework of the Lendület Programme.

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

Additional Information Both trials (CONSORT-PR and SPIRIT-PR) were prospectively registered on Open Science Framework (https://osf.io/c4hn8 and https://osf.io/z2hm9).

Differences Between Manuscript Versions: A Living Review and Series of Meta-Analyses

Mario Malički,[1,2,3] Ana Jerončić,[4] Gerben ter Riet,[5,6] Lex Bouter,[7,8] John P. A. Ioannidis,[2,3,9,10] IJsbrand Jan Aalbersberg,[11] Steven N. Goodman[1,2,3,9,12]

Objective Previous research has indicated a knowledge gap on changes brought on by peer review and journal publishing.[1] We are conducting a living evidence synthesis of studies that analyzed differences between manuscript versions (eg, preprinted or submitted versions vs peer reviewed journal versions).

Design This is a living synthesis of studies with meta-analyses of proportions (following PRISMA guidelines for reporting).[2] Studies were identified based on authors' knowledge of the field, from 18 previous systematic reviews

on peer review,[3] and by checking all research presented at the peer review conferences. All references and citations of identified studies (using Google Scholar up to January 2025) were checked. For all included studies that reported on changes, we extracted descriptive variables (including year of publication, sampling method, change comparison method, and data, wherever available, on how many analyzed version-pairs had changes for 11 outcomes). We assessed risk of bias for each outcome based on 3 domains: sample selection, measurement reliability, and conflicts of interest. To pool the data, we used random effects meta-analysis with Freeman-Tukey transformed proportions and reported the I^2 index of heterogeneity.

Results We identified 67 studies published from 1978 through the end of 2024, of which 31 (46%) analyzed changes between preprint and journal versions, 23 (34%) between submitted and accepted or published versions, 8 (12%) between rejected versions and those later published in other journals, 4 (6%) between multiple version sets, and 1 (1.5%) between different preprinted versions. Most studies, 45 (67%), analyzed changes manually, 13 (19%) used computational methods, and 9 (12%) combined manual and computational methods. The median number of analyzed version-pairs was 113 (IQR, 51-429). By discipline, 43 (64%) studies looked only at health research, 6 (9%) at life sciences, 6 (9%) at social sciences, 4 (6%) at physical sciences, and 8 (12%) at multiple disciplines. Meta-analyses showed the highest frequency of changes in title, authorship, conflicts of interest, and numerical results, with the lowest frequency in study conclusions (**Table 25-1077**).

Conclusions Current evidence indicates that while different manuscript sections experience varying degrees of changes from their submitted, preprinted, or rejected versions to their peer-reviewed journal version, the primary conclusions rarely change. Our results in this stage are limited by not conducting a systematic search of bibliographic databases and lack of (identified) data points from many subdisciplines, especially arts and humanities, as well as studies that explored possible factors associated with the extent of changes (eg, researchers' or editors' career stage or prestige, quality of initial manuscript version).

References

1. Tennant JP, Ross-Hellauer T. The limitations to our understanding of peer review. *Res Integr Peer Rev*. 2020;5:6. doi:10.1186/s41073-020-00092-1

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

3. European Association of Science Editors. What is peer review? December 28, 2024. Accessed July 15, 2025. https://ease.org.uk/communities/peer-review-committee/peer-review-toolkit/What-is-peer-review/

[1]Stanford Program on Research Rigor and Reproducibility (SPORR), Stanford University, Stanford, California, US, mmalicki@stanford.edu; [2]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, California, US; [3]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, US; [4]Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia; [5]Urban Vitality Centre of Expertise, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands; [6]Amsterdam University Medical Centers, Department of Cardiology, Amsterdam, The Netherlands; [7]Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, The Netherlands; [8]Amsterdam University Medical Centers, Department of Epidemiology and Data Science, Amsterdam, The Netherlands; [9]Department of Medicine, Stanford University School of Medicine, Stanford, California, US; [10]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, California, US; [11]Independent scholar; [12]Department of Health Policy, Stanford University School of Medicine, Stanford, California, US.

Conflict of Interest Disclosures Mario Malički was a co–editor in chief of Research Integrity and Peer Review, in which 1

Table 25-1077. Series of Meta-Analyses of Proportions on Changes Between Manuscript Versions

Change in	Summary proportion, % (95% CI)[a]	Heterogeneity, I^2, %	Version-pairs compared	Number of studies	Range of manuscript publication years
Title	24 (14-36)	99.7	155,403	15	1994 to 2023
Authorship	21 (17-25)	80.3	9046	10	1990 to 2023
Abstract conclusions[b]	6 (1-16)	87.4	345	3	2021 to 2022
Sample size	10 (6-15)	82.2	1191	11	2014 to 2023
Sample size calculation reporting	7 (3-12)	75.3	395	6	1990 to 2023
Numerical results[c]	31 (26-37)	74.7	1104	14	1996 to 2024
Results interpretation	7 (3-12)	82.0	628	8	2014 to 2024
Conclusions (discussion section)	2 (1-4)	56.6	1160	8	1996 to 2024
Limitations	19 (7-35)	94.5	578	7	1996 to 2024
Conflict of Interest declaration	26 (17-36)	77.8	430	5	2020 to 2024
Funding declaration	18 (9-28)	90.1	681	8	2012 to 2024

[a]Random effect meta-analyses with Freeman-Tukey arcsine square root transformation.

[b]Studies that looked only at how conclusions were phrased in the abstract.

[c]We grouped studies that reported any changes in any values reported or inclusion or removal of additional analyses. Note: Studies did not check whether added or removed numbers (pre-) existed in supplementary materials or were moved to them; they only noted changes in the main manuscript results section (including tables and figures).

of the included studies was published (he was not involved in the processing or decision making for that study, and Research Integrity and Peer Review full peer review histories are available to the public on the journal website). Mario Malički was also an external reviewer for 1 study. Gerben ter Riet is a co-author of 1 of the included studies and Steven N. Goodman is a co-author of 1 of the included studies. In both of these cases, data extractions and primary data analyses were done by Mario Malički and Ana Jerončić. IJsbrand Jan Aalbersberg was a Senior Vice President of Research Integrity at Elsevier. John P. A. Ioannidis, Lex Bouter, 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 The work on this project started in 2020, and in 2020 Elsevier funding was awarded to Stanford University for a 1-year METRICS postdoctoral position that supported Dr Malički's work. From 2021 to 2025, Dr Malički's work has been supported by the Stanford School of Medicine Research Office, which also supports the work of Dr Goodman. At the start of the project, IJsbrand Jan Aalbersberg was an employee of Elsevier.

Acknowledgments In 2022's Peer Review Congress, we presented preliminary findings (https://peerreviewcongress.org/abstract/a-synthesis-of-studies-on-changes-manuscripts-underwent-between-submission-or-preprint-posting-and-peer-reviewed-journal-publication/) that covered 25 studies published between 1990 and 2021. The current results include studies published before 1990 and up to January 2025.

Additional Information A website with computationally reproducible results from this living review will be available soon.

Quality of Patient Reviewer Comments and Association With Author and Editor Responses

Melecia Miller,[1] Vera Nezgovorova,[2] Ilana Kersh,[3] Mohamed I. Elsaid,[4] Marina Broitman[2]

Objective This study's goal was to develop a way to assess patient review quality based on author and editor reactions to reviews. We used patient reviews of draft final research reports (DFRRs) of Patient Centered Outcomes Research Institute (PCORI)–funded research that were peer reviewed by clinical, statistical, and patient reviewers.[1] Review quality ratings were based on prior research on PCORI peer review and other research on patient reviewers.[2,3] We hypothesized that specific, substantive changes to the report and patient-focused comments would be associated with a favorable mention of the patient review by the editor as well as a substantive response from the author.

Design Pilot testing on 13 patient reviews for 13 clinical research DFRRs submitted from 2021 to 2022 was completed in 2025. We conducted a content analysis of patient reviewer comments and developed a coding scheme based on prior research. Coding was performed at the sentence level and included 2 categories: (1) specificity and constructiveness of the comment and (2) patient focus of the comment (**Table 25-1110**). Each review was coded independently by 2 raters each for category 1 and category 2, with a final Cohen κ of 0.83 and 0.66, respectively, indicating moderate to substantial interrater agreement. Coding discrepancies between reviewers received the most conservative (ie, lowest value) code. Raters assessed whether the editor mentioned the patient review at least once in their review summary and either encouraged (code 1) or discouraged (code 2) an author response vs no mention (code 0). Author responses to reviewer comments were coded by paragraph indicating whether authors responded directly to patient-reviewer comments (code 1) or made changes to the report because of the comments (code 2) vs neither (code 0). We used hierarchical modeling to calculate the odds of editors encouraging authors to respond (code 1 vs codes 0 or 2) and mixed-effects logistic regression to calculate the odds of authors responding substantively (codes 1 or 2 vs code 0) to patient reviews with more specific, constructive comments.

Table 25-1110. Pilot Results Matrix for Category 1 and Category 2 Codes

Category 1: type of comment	Category 2: patient-focused comments, No. (%)			
	Code 0: comment is not patient focused (n = 122)	Code 1: comment focused on patient engagement (n = 127)	Code 2: comment focused on patient centeredness (n = 222)	Total comments by type (N = 471)
Comment suggests specific, substantive changes	40 (32.79)	31 (24.41)	64 (28.83)	135
Comment suggests vague criticism or changes	20 (16.39)	9 (7.09)	42 (18.92)	71
Comment provides only praise for the research	18 (14.75)	28 (22.05)	31 (13.96)	77
Neutral observation, summary, or grammatical change	44 (36.07)	59 (46.46)	85 (38.29)	188

Results Table 25-1110 shows counts for associations between category 1 and category 2. More specific and constructive comments and patient-focused comments were not associated with higher odds of the editor mentioning the patient review (N = 13; odds ratio [OR], 0.36; 95% CI, 0.28-0.45 and OR, 0.65; 95% CI 0.54-0.78, respectively). Patient review paragraphs with more than 50% specific and constructive comments were associated with higher odds of authors responding substantively, compared with other comment types (N = 147; OR, 4.00; 95% CI, 1.67-9.59; *P* = .002).

Conclusions This pilot study shows the feasibility of establishing codes rating the quality of patient reviews and determining whether higher quality patient reviews are associated with mentions by the editor and author responses. The results are limited because of the sample size and the uniqueness of DFRRs.

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. Ghazaryan E, Broitman M, Sox H. Differences in the style and quantity of reviewer comments in structured vs unstructured peer review forms. Poster presented at: Ninth International Congress on Peer Review and Scientific Publication; September 10, 2022; Chicago, IL.

3. Huml AM, Albert JM, Beltran JM, et al. Community members as reviewers of medical journal manuscripts: a randomized controlled trial. *J Gen Intern Med*. 2022;38(6):1393-401. doi:10.1007/s11606-022-07802-z

[1]KnowledgeWorks Global, Ltd., 410 MN-25, Brainerd, MN, US; mmiller@originreview.org; [2]Patient-Centered Outcomes Research Institute, Washington, DC, US; [3]University of Pittsburgh, Pittsburgh, PA, US; [4]The Ohio State University College of Medicine, Columbus, Ohio, US.

Conflict of Interest Disclosures None reported.

Factors Associated With Outcomes of Appeals of Manuscripts Initially Rejected by a General Medical Journal

Matthew B. Stanbrook,[1] Shannon Charlebois,[1] George A. Tomlinson,[1] Meredith Weinhold[1]

Objective The opportunity to appeal a journal's decision to reject a manuscript enhances procedural fairness toward authors and allows editors to correct errors in judgment but wastes time for both authors and editors when unsuccessful. We sought to identify factors associated with outcomes of appeals of research manuscripts at the *Canadian Medical Association Journal* (*CMAJ*), a general medical journal that explicitly highlights the option to appeal in rejection letters.

Design We conducted a retrospective cohort study of all appeal requests for manuscripts initially rejected after submission to the Research section of *CMAJ* between 2007 and 2024. We obtained data on manuscript characteristics and editorial decisions from *CMAJ*'s manuscript submission database (ScholarOne). Three editors independently reviewed the text of the initial manuscript decision letters to identify reasons for rejection, with disagreements resolved by consensus. We included manuscripts rejected at any stage of the editorial process. The primary outcome was whether appeal requests were granted or denied. A secondary outcome was whether appealed manuscripts were ultimately published. We estimated associations of multiple prespecified characteristics with appeal outcomes using bayesian logistic regression.

Results Among 14,351 submitted manuscripts, 13,419 were rejected, and 616 of these rejections (4.6%) were appealed. Among appeals, 214 (34.7%) were granted, of which 58 (27.1% of granted appeals and 9.4% of all appeals) were ultimately published. Compared with manuscripts rejected without external peer review, granted appeals were more likely for manuscripts rejected after peer review without editorial discussion (adjusted odds ratio [AOR], 2.71; 95% credible interval [CrI], 1.40-5.27) but similarly likely for manuscripts rejected after peer review and editorial discussion (AOR, 1.19; 95% CrI, 0.74-1.91) (**Table 25-1161**). Granted appeals were more likely if rejection letters included positive comments (AOR, 1.82; 95% CrI, 1.21-2.76) or cited lack of clarity (AOR, 1.84; 95% CrI, 1.12-3.04) and less likely if corresponding authors were based outside Canada (AOR, 0.38; 95% CrI, 0.24-0.59). The odds of an appeal being granted also varied significantly based on the individual editor making the appeal decision (median AOR, 1.90; 95% CrI, 1.24-4.37). Among factors associated with appeals being granted, only positive comments (AOR, 1.98; 95% CrI, 1.10-3.59) were also significantly associated with successful publication after appeal. However, appealed manuscripts rejected after peer review and editorial discussion were more likely to be published than were those rejected without peer review (AOR, 4.93; 95% CrI, 2.47-10.25).

Conclusions Characteristics of authors, editors, editorial process, and decision communication were associated with successful appeals of initially rejected research manuscripts. Lack of awareness of the small fraction of appeals that ultimately achieve publication may raise false expectations among authors.

Table 25-1161. Adjusted ORs for Appeals of Manuscript Rejection Decisions Being Granted vs Denied for Selected Characteristics

Variable	Appeals, No. (%)[a]		Adjusted OR (95% CrI)[b]
	Granted (n = 214)	Denied (n = 402)	
Editorial stage at original rejection			
Without external peer review	95 (28.1)	239 (70.7)	1 [Reference]
After peer review without editorial discussion	30 (44.1)	38 (55.9)	2.71 (1.40-5.27)
After peer review and editorial discussion	85 (42.3)	116 (57.7)	1.19 (0.74-1.91)
After revision	4 (44.4)	5 (55.6)	1.19 (0.23-5.94)
Corresponding author based outside Canada	43 (20.1)	152 (37.8)	0.38 (0.24-0.59)
Rejection letter word count quartilec			
52-76	NA	NA	0.95 (0.49-1.83)
77-105	NA	NA	1.18 (0.60-2.28)
106-180	NA	NA	1.22 (0.60-2.45)
181-1285	NA	NA	2.00 (0.90-4.48)
Rejection letter included positive comments	119 (55.6)	152 (37.8)	1.82 (1.21-2.76)
Rejection letter cited lack of clarity	58 (27.1)	57 (14.2)	1.84 (1.12-3.04)

Abbreviations: CrI, credible interval; NA, not applicable; OR, odds ratio.

[a]Row percentages are shown.

[b]Adjusted ORs were calculated using bayesian logistic regression.

[c]The Canadian Medical Association Journal's rejection letter templates include a specific section for editors to explain the reasons for a manuscript decision. Word counts are specific to this section of the letter only.

[1]*Canadian Medical Association Journal* (*CMAJ*), Toronto, Ontario, Canada, matthew.stanbrook@cmaj.ca.

Conflict of Interest Disclosures All authors are employees or paid consultants of *CMAJ*. Matthew B. Stanbrook, Shannon Charlebois, and George A. Tomlinson participated in editorial decision making for some of the manuscripts included in the study. No other disclosures were reported.

Automated Targeted E-mails for Improving Author Compliance With Study Reporting Requirements and Other Editorial Processes

Daniel S. Evanko,[1] Deondre Jordan[1]

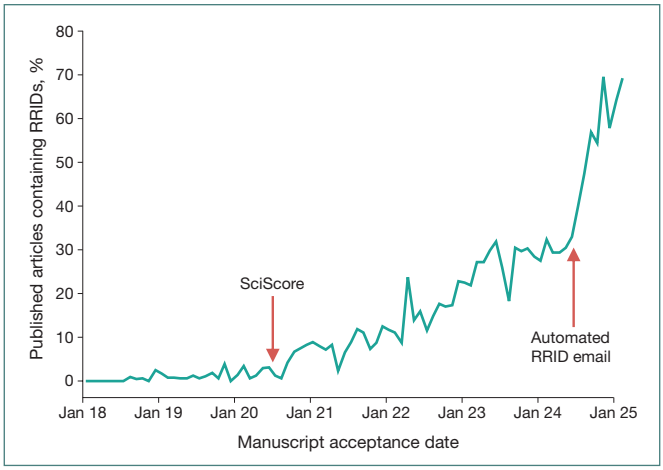
Objective Journals ask authors to comply with many reporting requirements to improve study transparency, reproducibility, and reuse. Manual monitoring and enforcement of requirements is difficult, and reporting checklists or vendor-based solutions can be untargeted, burdensome, expensive, or of limited effectiveness. For example, since June 2020 the American Association for Cancer Research has used a methods review tool (SciScore) to help improve author compliance with reporting requirements at its 10 journals. This tool greatly increased author assignment of research resource identifiers (RRIDs),[1] but compliance plateaued at approximately 30% of original research articles published in the association journals. This cross-sectional study describes a simple, inexpensive, automated approach to further overcome limitations of common approaches and quantitatively evaluates its performance using 3 example interventions.

Design Custom Python scripts coupled with structured language queries were used for targeted extraction of manuscript data from the EJournal Press journal submission system. For RRID reporting compliance, once every day the submission system was queried for manuscripts that completed initial quality control checks and did not contain the text RRID. A templated e-mail alert was sent to the corresponding author that explained what RRIDs are, why they are important, and how to add them. The system was also adapted with a custom supervised machine-learning method to identify studies that generated new sequencing data or that were computationally intensive to encourage authors to deposit sequence data and provide computer code used in their study and used to send different alerts for these manuscripts at different designated points in the editorial process.

Results The new automated RRID process was activated August 6, 2024, at all the association journals, resulting in missing-RRID alerts sent for 84% (4949 of 5899) of manuscripts at completion of initial quality control checks through the end of 2024. Performance was assessed by searching for "RRID" in published articles as done previously.[1] An immediate sharp increase in articles containing RRIDs was observed that quickly increased the percentage of published research articles containing RRIDs from 33% (176 of 526) in April through July to 65% (330 of 505) in November through February **(Figure 25-1181)**. Targeted alerts to authors of manuscripts asking them to deposit sequencing data (sent after peer review started) or supply code in a computational science software (Code Ocean) capsule (sent after a revise decision was sent to the author) were turned on on August 26, 2024, and resulted in sequence deposition alerts sent for 43% (1416 of 3261) of manuscripts and Code Ocean alerts sent for 39% (614 of 1594) of manuscripts through April 30, 2025. The sequencing alert triggered more than 100 author e-mail communications about deposition, demonstrating author engagement with this request and allowing resolution early in the editorial process. The percentage of accepted manuscripts with linked Code Ocean capsules increased from 1.4% (21 of 1555) in 2024 to 4.6% (25 of 547) in the January to May 2025 period, a more than 3-fold increase.

Conclusions This fully automatic and highly targeted process engages with authors in a context-dependent way at predefined times in the editorial process that are appropriate for the specific communication, thus achieving substantial author engagement and improved compliance with journal requirements.

Figure 25-1181. Prevalence of Research Resource Identifiers (RRIDS) in Published Original Research Manuscripts Based on the Month They Were Accepted



Arrows indicate the time that the indicated systems were activated.

Reference

1. Roelandse M, Ozykurt IB, Evanko D, Bandrowski A. Assessing the effectiveness of SciScore in supporting the reproducibility of scientific research. *Sci Ed.* 2023;46:46-52.

[1]American Association for Cancer Research, Philadelphia, PA, US, 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.

Virtual

Trends in Peer Review Metrics at the *Annals of African Surgery*

James Kiilu,[1] Cecilia Munguti,[1] James Kigera,[1] Michael Mwachiro[1]

Objective The peer review process faces the challenges of recruiting reviewers and assessing the quality of the peer review process and manuscripts.[1] This study evaluated trends in reviewer performance (R-scores), manuscript quality (M-scores), and their association with reviewer recommendations and editorial decisions at the *Annals of African Surgery* over 7 years.

Design This retrospective study analyzed peer review reports at the *Annals of African Surgery* from 2017 to 2023, using data extracted from the journal's manuscript management system. Variables included manuscript ID, submission year, revision cycles, reviewer ID, recommendations, R-scores, M-scores, time to review, and editorial decisions. R-scores assessed review timeliness and relevance, while M-scores evaluated the manuscript across dimensions such as interest, quality, and originality. Descriptive statistics summarized reviewer invitation responses, and cross-tabulations were performed to examine trends in R-scores and M-scores against submission year, reviewer recommendations, revision cycles, and editorial decisions. Associations between reviewer recommendations and editorial outcomes were assessed using χ^2 tests. Continuous data were tested for skewness, with analysis of variance applied to normally distributed data and Kruskal-Wallis tests to skewed data.

Results For the period under review, the journal had a total of 1193 reviewers, with 370 (31%) coming from the international community. The number of reviewers invited per manuscript varied significantly across years, as determined by an independent-samples Kruskal-Wallis test (H_2 = 18.46; *df* = 6; *P* = .005). The mean (SD) number of reviewers invited per manuscript was 6.35 (3.4) in 2017, peaked at 9.81 (9.6) in 2019, then declined to 5.12 (3.5) in 2021 before rising again to 7.02 (4.03) in 2023. There was a significant improvement in the reviewer acceptance rate, increasing from 29.9% in 2019 to 45.7% in 2023 (*P* < .001) (**Table 25-0922**). R-scores decreased significantly over the study period, from mean (SD) 2.68 (0.53) in 2017 to 2.29 (0.65) in 2023 (*P* < .001), while there was no significant change in M-scores. There was no statistically significant change in R-scores across different manuscript revision cycles, but M-scores showed significant improvement (from mean [SD] 2.60 [0.84] in the initial round of review to 2.17 [0.61] after 3 revisions, *F* = 2.413; *P* < .05), accompanied by a marked reduction in review turnaround time (from mean [SD] 11.07 [9.9] to 1.29 [1.38] days; *P* = .01). R-scores and M-scores varied significantly by editorial decision, with the highest R-scores observed for manuscripts receiving a major revision decision (mean = 2.71; *P* < .001) and the highest M-scores seen in rejected manuscripts (mean = 3.91; *P* < .001). One-way analysis of variance demonstrated statistically significant differences in mean R-scores awarded by the 15 editors ($F_{14, 3986}$ = 17.77; *P* < .001). Reviewer recommendations were associated with editorial decisions (Pearson χ^2 = 288.80; *df* = 9; N = 1173; *P* < .001).

Table 25-0922. Trends in Reviewer Acceptance, R-Scores, M-Scores, and Time to Review

Variable (Unit)	Year							*P* value
	2017	2018	2019	2020	2021	2022	2023	
No. of invited reviewers per year, mean	6.35	8.13	9.81	6.28	5.12	5.78	7.02	.005
Reviewer acceptance rate, %[a]	57.1	28.4	19.7	25.2	28.8	30.0	45.7	<.001
R-score, mean, 0–3[b]	2.68	2.57	2.38	2.42	2.48	2.3	2.29	<.001
M-score, mean, 0–5[c]	2.6	2.48	2.59	2.62	2.58	2.53	2.64	.58
Time to review, d[d]	10.03	10.25	11.12	10.62	11.01	10.64	11.97	.14

[a]Reviewer acceptance rate: percentage of invited reviewers who accepted the invitation to review.

[b]R-score: reviewer performance score assessing timeliness and relevance (higher scores indicate better performance).

[c]M-score: manuscript quality score (higher scores indicate lower quality).

[d]Time to review: average number of days from review assignment to completion.

Conclusions While reviewer acceptance rates improved significantly over time, there was no corresponding improvement in R-scores or M-scores. Future efforts should prioritize not only increasing reviewer participation but also strengthening the quality and consistency of reviewer feedback to more effectively support manuscript development.

Reference

1. Peterson CJ, Orticio C, Nugent K. The challenge of recruiting peer reviewers from one medical journal's perspective. *Proc Bayl Univ Med Cen*t. 2022;35(3):394-396. doi:10.1080/08998280.2022.2035189

[1]Annals of African Surgery, Nairobi, Kenya; james.kiilu@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.

Impact of a Novel Checklist on the Peer Review Process

Jorge Finke,[1] Sumi Sexton[2]

Objective The peer review process has been criticized for a lack of evidence that it improves the quality of scientific literature.[1] Some authors have reported peer reviews to be incoherent, lacking constructive feedback, and slow.[2] Importantly, little research has been done on peer review for narrative reviews. We have developed a checklist for peer reviewers to evaluate the effects on the quality of peer reviews and manuscripts.

Methods This was a pseudorandomized nonblinded clinical trial. A 10-question checklist was developed for use by peer reviewers evaluating manuscripts for the medical journal *American Family Physician (AFP)*. The checklist was created through the Delphi method, utilizing a panel of experts including *AFP* medical editors and editors from other journals who reviewed the proposed checklist. The checklist was internally validated by 2 medical editors independently comparing results of the new checklist to the journal's existing manuscript quality scoring system. Manuscripts were assigned on alternating months to either the experimental group in which reviewers and editors applied the new checklist or the control group which utilized the traditional system. After peer review, the medical editor assigned to each manuscript rated the quality of each review, using a 0 to 100% scale to determine whether the checklist improved peer reviewer performance. Quality ratings of the initially submitted and finalized manuscripts were also determined by

the medical editor using a 0 to100% scale. Changes in initial to finalized manuscript quality ratings were compared between the experimental and control groups.

Results A total of 78 manuscripts were evaluated from December 2022 through March 2024. Forty of these manuscripts were assigned to peer reviewers utilizing the newly developed checklist, and 38 used the traditional review process. No statistically significant improvement was noted in review quality (68.5%; 95% CI, 63.5%-72.8% vs 65.7%; 95% CI, 61.0%-69.8%; *P* = .40), review word count length (683 words; 95% CI, 608.8-754.5 words vs 616.04 words; 95% CI, 545.8-684.1 words; *P* = .19) or speed of the peer review process (45.49 days; 95% CI, 37.2-52.8 days vs 37.3 days; 95% CI, 31.9-41.6 days; *P* = .08). Similarly, no statistically significant improvement was found in manuscript quality rating after revisions based on peer review with new checklist compared with the control group (difference, 15.7%; 95% CI, 9.1%-22.6% vs 12.6%; 95%CI, 6.7%-18.8%; *P* = .49).

Conclusions Few interventions have had positive impacts in the peer-review process, including the appropriate use of checklists or guidelines.[3] While this study showed no significant improvements with the new checklist in the peer review quality rating, reviewer word counts, speed of reviewer process, and overall manuscript ratings, no parameters were negatively affected. The checklist was still formally adopted as a modified version of the previous format developed based on Delphi feedback from experts in the field.

References

1. Kelly J, Sadeghieh T, Adeli K. Peer review in scientific publications: benefits, critiques, & a survival guide. *EJIFCC*. 2014;25(3):227-243.

2. Sciullo N, Duncan M. Professionalizing peer review suggestions for a more ethical and pedagogical review process. *J Schol Pub*. 2019;50:248-264. doi:10.3138/jsp.50.4.02

3. Gaudino M, Robinson NB, Di Franco A, et al. Effects of experimental interventions to improve the biomedical peer-review process: a systematic review and meta-analysis. *J Am Heart Assoc*. 2021;10(15):e019903. doi:10.1161/JAHA.120.019903

[1]Contributing editor, *American Family Physician*, Leawood, KS, US, jfinke@bidmc.harvard.edu; [2]Editor in chief, *American Family Physician*, Leawood, KS, US.

Conflict of Interest Disclosures None reported.

Acknowledgment This study was conducted with support from *American Family Physician* as an internal quality improvement project and the intervention developed as adopted for use in the journal's peer review process. This work was conducted with support from UL1TR002541 award through Harvard Catalyst (Biostatistics/Bioinformatics Consultation Program), The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health), and financial contributions from Harvard University and its affiliated academic health care centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the National Institutes of Health.

Identifying Methodological Concerns in Agency for Healthcare Research and Quality Evidence-Based Practice Center Reports: Analysis of Editorial Review Comments

Haley K. Holmer,[1] Edi E. Kuhn,[1] Camber Hansen-Karr,[1] Ed Reid,[1] Mark Helfand[1]

Objective To identify methodological concerns in Agency for Healthcare Research and Quality (AHRQ) Evidence-Based Practice Center (EPC) Program reports and assess the comprehensiveness of EPC Methods Guidance using the Systematic Reviewlution (SR) Framework.[1]

Design Cross-sectional analysis of reviewer comments from 25 EPC reports submitted between June 1, 2018, and March 1, 2023. The sample included systematic reviews (n = 16), systematic review updates (n = 4), and technical briefs (n = 5). For the internal, pre–peer-reviewed phase, 1 reviewer extracted comments received from associate editors, task order officers, and AHRQ leadership. In the external peer-reviewed phase, we extracted comments from peer reviewers, key informants, technical experts, partners, and public reviewers. A senior reviewer used the SR Framework,[1] a comprehensive typology of 67 systematic review problems, to classify concerns raised in the pre–peer-reviewed and/or peer-reviewed phase and to quantify frequencies. We identified patterns in 2 areas: (1) concerns raised in the pre–peer-review phase that persisted into the peer-reviewed phase and (2) concerns missed in the pre–peer-reviewed phase that were raised in the peer-reviewed phase. Further, we determined whether problems identified in the SR Framework are addressed in EPC Methods Guidance,[2,3] and we identified methodological concerns absent from the SR Framework.

Results Of 5024 total reviewer comments, 15% (252 of 1717) of pre–peer-reviewed comments and 10% (348 of 3307) of peer-reviewed comments addressed methodological concerns. Nearly half of the methodological concerns identified during the pre–peer-review phase pertained to grading strength of evidence (SOE) (n = 116 [46%]), followed by concerns with assessing risk of bias (n = 23 [9%]), and spin (n = 21 [8%]). The most frequent concerns raised by peer reviewers also related to spin (n = 56 [16%]), grading SOE (n = 44 [13%]), and lack of clinical expert or stakeholder perspective (n = 41 [12%]). Spin-related concerns included conclusions that went beyond the evidence, policy implications disconnected from results, failure to incorporate SOE into conclusions, and biased tone in presentation. Comments raised in the peer-reviewed phase that were not raised in pre–peer-review phase included concerns about a priori protocol decisions or inflexible methods for answering review questions. The SR Framework failed to capture several key methodologic issues identified in EPC reports, including concerns about (1) SOE ratings, (2) inadequately defined outcomes, and (3) a priori

protocol registration. Over 75% (n = 51) of the 67 problems in the SR Framework were addressed in EPC methods guidance.

Conclusions Overall, EPC reports are of high methodological quality. Notable patterns in methodological concerns emerged, with specific issues pertaining to grading SOE, spin, need for clinical expert or stakeholder perspectives, and balancing adherence to an a priori protocol against potential bias introduced by an adaptive protocol. The persistence of spin-related concerns across both review phases suggests a challenge in maintaining objectivity in evidence synthesis. Notably, the SR Framework did not capture important methodological concerns in EPC reports, suggesting potential gaps in the SR Framework for evaluating the trustworthiness of evidence synthesis products.

References

1. Uttley L, Quintana DS, Montgomery P, et al. The problems with systematic reviews: a living systematic review. *J Clin Epidemiol*. 2023;156:30-41. doi:10.1016/j.jclinepi.2023.01.011

2. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Page last reviewed October 2022. Effective Health Care Program, Agency for Healthcare Research and Quality. https://effectivehealthcare.ahrq.gov/products/collections/cer-methods-guide

3. 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

[1]Portland VA Research Foundation, Portland, OR, haley.holmer@va.gov.

Conflict of Interest Disclosures None of the authors have any affiliations or financial involvement that conflicts with the material presented in this abstract.

Funding/Support This work is funded by the Agency for Healthcare Research and Quality (AHRQ) Effective Healthcare (EHC) Program through a contract to the Scientific Resource Center (Contract No. HHSA 75Q80122C00002). 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 AHRQ, US Department of Health and Human Services.

Education/Training

In-person

Peer Review Exercises to Enhance Trainees' Readiness to Confront Unfair or Biased Reviews

Franki Y. H. Kung,[1] Mariam Aly,[2] Shahana Ansari,[3] Eliana Colunga,[4] M. J. Crockett,[5,6] Amanda B. Diekman,[3] Pablo Gomez,[7] Paul C. McKee,[8] Miriam Pérez,[9] Sarah M. Stilwell,[10] and Matthew Goldrick[11,12]

Objective Accumulating data suggest peer reviews can exhibit bias (ie, systematically devaluing certain topics, methods, or the study of nondominant populations[1]) and can be unfair with respect to process (eg, decisions are made inconsistently) or people (eg, authors are disrespected).[2] Recent studies have shown that biases and unfair behaviors in peer review tend to undermine the self-reported experiences and career progression of researchers, including those from underserved populations, such as women, nonbinary individuals, and people from racial and ethnic groups other than White.[1-3] This study examined whether peer review exercises increased psychology and neuroscience trainees' perceived readiness to respond to biased and unfair reviewers.

Design Single-session synchronous online trainings (conducted during 2023-2024) allowed for trainer-trainee and trainee-trainee interactions. Trainees were recruited for this study through social media, targeted academic organizations, and graduate programs. Trainers (professors and a postdoctoral fellow with multiple years of peer review experience) reviewed processes occurring between journal submission and receipt of a decision letter. Trainees were then given a simplified, anonymized decision letter and reviews. Exercises and discussions illustrated how trainees could plan manuscript revisions and compose responses to critiques. Finally, trainees were given examples of biased and unfair reviews, along with strategies for responding to them while acknowledging the power dynamics inherent to review. Discussion included how to maintain motivation when encountering such reviews. Participants were asked to complete surveys before and after online training in which they rated their ability to perform the following targeted skills on a scale from 1 (very little skill) to 5 (great deal of skill): define and identify bias in peer reviews; define and identify unfair language in peer reviews; respond to reviews using fair language; respond to biased or unfair reviews; and promote an inclusive and equitable peer-review culture.

Results A total of 64 participants (42 PhD students [66%], 10 predoctoral students [16%], 8 postdoctoral trainees [13%], and 4 other trainees [5%]) rated their level of skill for 5 targeted skills. Ordinal regressions showed participants had significantly higher ratings for all skills at posttest. For example, the mean (SD) self-ratings for the skill to respond to biased on unfair reviews increased from 2.0 (1.01) on the pretest to 4.09 (0.89) on the posttest (**Table 25-0893**).

Conclusions After these peer review exercises, psychology and neuroscience trainees had an increase in their perceived readiness to confront biased and unfair reviews. Work is ongoing to scale up this training using on-demand materials and further assess training impact, including measures beyond trainee self-reports.

References

1. Rogers LO, Moffitt U, McLean KC, Syed M. Research as resistance: naming and dismantling the master narrative of "good" science. *Am Psychol*. 2024;79:484-496. doi:10.1037/amp0001246

2. Silbiger NJ, Stubler AD. Unprofessional peer reviews disproportionately harm underrepresented groups in STEM. *PeerJ*. 2019;7:e8247. doi:10.7717/peerj.8247

Table 25-0893. Mean Pretest vs Posttest Self-Ratings of Peer Review Skill With Ordinal Regression Coefficient and Wald *z* Statistic[a]

Skill	Pretest (SD)	Posttest (SD)	Posttest vs pretest coefficient (SE)	Wald *z*
Define and identify bias in peer reviews	2.3 (1.05)	4.06 (0.77)	3.37 (0.45)	7.42
Define and identify unfair language in peer reviews	2.5 (1.28)	4.17 (0.73)	2.76 (0.41)	6.79
Respond to reviews using fair language	2.69 (1.21)	4.16 (0.86)	2.48 (0.39)	6.38
Respond to biased or unfair reviews	2.0 (1.01)	4.09 (0.89)	3.91 (0.50)	7.74
Promote an inclusive and equitable peer-review culture	2.7 (1.18)	4.08 (0.93)	2.27 (0.37)	6.09

[a]Skills were rated on a scale of 1 (very little skill) to 5 (great deal of skill).

3. Aly M, Colunga E, Crockett MJ, et al. Changing the culture of peer review for a more inclusive and equitable psychological science. *J Exp Psychol Gen*. 2023;152:3546-3565. doi:10.1037/xge0001461

[1]Department of Psychological Sciences, Purdue University, West Lafayette, IN, US; [2]Department of Psychology, University of California Berkeley, Berkeley, CA, US; [3]Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, US; [4]Department of Psychology and Neuroscience, University of Colorado Boulder, City, CO, US; [5]Department of Psychology, Princeton University, Princeton, NJ, US; [6]Department of Psychology, University Center for Human Values, Princeton University, Princeton, NJ, US; [7]Department of Psychology, Skidmore College, Saratoga Springs, NY, US; [8]Department of Psychology and Neuroscience, Duke University, Durham, NC, US; [9]Department of Psychology, North Park University, Chicago, IL, US; [10]Department of Health Behavior and Health Equity, University of Michigan, Ann Arbor, MI, US; [11]Department of Linguistics, Northwestern University, Evanston, IL, US, matt-goldrick@northwestern.edu; [12]Cognitive Science Program, Northwestern University, Evanston, IL, US.

Conflict of Interest Disclosures None reported.

Funding/Support This work was supported by grants from the National Science Foundation (DGE-2224777, DGE-2224779, DGE-2436430).

Role of the Funder/Sponsor The funder of the study 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.

Additional Information Training materials are freely available (https://osf.io/658ev).

Virtual

A Pilot Program for Early Career Mentorship in Journal Peer Review

Anjali Garg,[1] Preeti Panda,[2] Lydia Furman,[3] Kimberly Montez,[4] Alex R. Kemper,[5] Lewis First[6]

Objective To implement and evaluate a pilot mentorship program in which experienced peer reviewers mentor novice reviewers in writing peer reviews.

Design In 2023, editorial board members of *Pediatrics*, the peer-reviewed journal of the American Academy of Pediatrics, volunteered for a pilot mentorship program to improve the peer review skills of early career physicians (medical students to junior faculty) to enhance the number of early career reviews within the pool present at *Pediatrics*. Mentors selected 1 or 2 mentees. The target participation for the pilot was 10 dyads. Literature review and expert consensus informed development of peer review mentorship materials. Curricula for mentees included didactics and checklists, and material for mentors provided feedback guidance. For objective assessment, the review quality instrument (RQI) was adapted into an 8-item, 3-point Likert scale questionnaire to score mentee reviews.[1] Mentees needed to complete at least 3 reviews with a score of 3 on at least 6 of 8 RQI items to graduate the program—a metric selected by the program leaders. Mentors and mentees completed a review of a regular article (as defined by the *Pediatrics* author guidelines) separately, then met together for structured feedback from the mentor to the mentee on the review. Each regular article was sent to the mentor based on their area of expertise; the turnaround time for completion was 3 weeks. The RQI was completed by the mentor, and the combined review was ultimately submitted to the editorial board by the mentor. Qualitative data regarding skills gained and areas for improvement were compiled and analyzed from intake and exit surveys completed by mentors and mentees.

Results Eleven dyads enrolled, with 9 graduating from the program, an 81% graduation rate. Five dyads (55%) required 3 reviews to graduate, 3 (33%) required 4 reviews, and 1 (9%) required 5 reviews. The median number of RQI items scoring a 3 in each dyad's first reviews increased from 4 (IQR, 2-7) to 7 (IQR, 6-8) by the last review. The range of total RQI scores increased from 16-24 (first reviews) to 23-24 (last reviews) for each dyad. Qualitative analysis indicated that the pilot program was feasible and effective. Notably, mentees reported increased comfort with peer review and improvement in their own manuscript writing skills. Mentors noted an improvement in their own reviewing skills through mentorship. Areas for improvement included targeted communication from the leadership team and allowing separate enrollment of mentees and mentors with matching by the program team to further enhance dissemination of this program.

Conclusions The pilot program was successful in enrolling the target number of participants, graduating over 80% of the mentees with improvement of reviewing skills per increase in RQI scores, and demonstrated feasibility for future program building.

Reference

1. 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

[1]Division of Critical Care Medicine, Department of Pediatrics, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, US, anjaliga07@gmail.com; [2]Division of Pediatrics, Department of Emergency Medicine, Stanford University, Stanford, CA, US; [3]University Hospitals Rainbow Babies & Children's, Cleveland, OH, US; [4]Department of Pediatrics, Wake Forest University School of Medicine, Winston-Salem, NC, US; [5]Division of Primary Care Pediatrics, Nationwide Children's Hospital, Columbus, OH, US; [6]Department of Pediatrics, University of Vermont Children's Hospital, Burlington, VT, US.

Conflict of Interest Disclosures All authors are members of the *Pediatrics* editorial board.

Acknowledgment We acknowledge Kate Larson and the *Pediatrics* editorial board administrative team for their help with the implementation of this program.

Additional Information Preeti Panda is a co–corresponding author (preetipanda@stanford.edu).

In-Person Peer Review Training to Improve Preparedness to Evaluate Manuscripts

Marcus Tolentino Silva,[1] Taís Freire Galvão[2,3]

Objective Peer review is the core of scientific publication, and mentoring seems to enhance the quality of reviews.[1] We aimed to assess the effect of peer review training on reviewers' preparedness to evaluate manuscripts.

Design This study was designed to improve peer review quality for a diamond open access journal, *Epidemiologia e Serviços de Saúde: revista do SUS* (RESS). The intervention consisted of a 16-hour, in-person, 2-day workshop.[2] A checklist consisting of 31 critical, important, or desirable items for the preparation of peer review opinions was developed, and its use was encouraged to guide reports.[3] The courses took place in June (including associated editors and reviewers), September, and October 2024 in 3 Brazilian cities and consisted of practical sections of peer review, preceded by an explanation of the checklist and peer review process. Instructors were available for students' questions, and discussions about bias in scientific communication were stimulated. Before starting, participants were invited to answer an anonymized questionnaire to assess their reviewing skills by means of a 5-point Likert scale (1: poor, 5: excellent), and another survey was sent electronically in January 2025 to assess checklist usefulness and the mean difference (MD) and 95% confidence interval (95% CI) of pre- and post-event confidence as a peer reviewer. We described participants' characteristics and summarized means and standard deviations (SDs), and tested differences in distribution using Fisher exact, analysis of variance, Kruskal-Wallis, or *t* tests. All analyses were performed in Stata v.14.2. The project was approved as an extension activity of the University of Brasília, where all participants registered themselves.

Results In total, 96 participants were trained; 51 (76.1%) were women and 38 (56.7%) had a doctoral degree, which was more frequent in first-edition participants (26 [70.3%]; $P = .02$). For 67 initial survey respondents, the mean (SD) familiarity with article submission was 4.1 (1.1) and with peer review, 3.5 (1.4). Mean (SD) confidence as a peer reviewer was 3.5 (1.3). The January 2025 survey had 43 respondents, who reported having issued a mean (SD) of 7.6 (25.9) peer reviews since the workshops. Confidence as a peer reviewer significantly increased (MD, 0.5 [95% CI, 0.1 to 1.0]; $P = .01$); 86.1% (n = 37) reported using the checklist to issue peer reviews and gave it a mean (SD) rating of 4.6 (0.9). Results are summarized in **Table 25-0888**.

Conclusions Peer review practical training sections and a checklist with suggested items improved peer reviewers' confidence in evaluating manuscripts. Taking into account the costs and sustainability issues involved, this strategy may be viable to improve peer reviews.

References

1. Lyons-Warren AM, Aamodt WW, Pieper KM, Strowd RE. A structured, journal-led peer-review mentoring program enhances peer review training. *Res Integr Peer Rev*. 2024;9(1):3. doi:10.1186/s41073-024-00143-x

2. Silva MT, Galvão TF. Systematization of peer review in *Epidemiologia e Serviços de Saúde*. *Epidemiol Serv Saude*. 2024;33:e20241001. doi:10.1590/S2237-96222024v33e20241001.en

3. Silva MT, Galvão TF. Revisão por pares: itens recomendados na elaboração de pareceres [Peer review: recommended items for preparing opinions]. June 19, 2024. Accessed May 24, 2025. https://osf.io/grn2a

[1]Faculdade de Ciências da Saúde, Universidade de Brasília, Brasília, Brazil; [2]Faculdade de Ciências Farmacêuticas, Universidade Estadual de Campinas, Campinas, Brazil, taisgalvao@gmail.com; [3]Secretaria de Vigilância em Saúde e Ambiente, Ministério da Saúde, Brasília, Brazil.

Conflict of Interest Disclosures Taís Freire Galvão reports being an executive editor of *Epidemiologia e Serviços de Saúde: revista do SUS* (RESS). Taís Freire Galvão 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.

Quality of an Educational Program to Empower Early Career Faculty and Trainees Through Mentored Training in Peer Review

Susan Galandiuk,[1] Vaitheesh Jaganathan,[2] Hillary Simon[1]

Objective Despite advances in electronic manuscript submission systems, the editorial peer review process for medical journals has had few fundamental changes aside from promptness of review. In 2018, the editor of a society-owned journal created the *Reviewer's Guild*, a mentored 1-year peer-review training program designed for surgical trainees, fellows in specialty training, and young academic faculty. The goal of the program was to attract the brightest individuals in the field, elevate the quality of peer review, and create a feeling of loyalty toward the journal. The program consists of a formal curriculum followed by 5 mentored

Table 25-0888. Participants' Characteristics and Skills in Peer Review

Variable	1st course (n = 51)	2nd course (n = 20)	3rd course (n = 25)	All courses (N = 96)	*P* value
Initial survey respondents, No.	37	9	21	67	.02[a]
Women, No. (%)	27 (72.9)	8 (88.9)	16 (76.2)	51 (76.1)	.73[a]
Age in years, mean (SD)	42.4 (8.2)	32.7 (7.0)	42.6 (6.7)	41.2 (8.2)	.003[b]
Doctoral degree, No. (%)	26 (70.3)	2 (22.2)	10 (47.6)	38 (56.7)	.02[a]
Time since graduation in years, mean (SD)	19.1 (8.4)	6.1 (7.0)	20.0 (6.0)	17.7 (8.7)	<.001[b]
No. of articles read in the previous week, mean (SD)	4.4 (3.1)	5.1 (6.1)	4.3 (6.3)	4.5 (4.6)	.91[b]
No. of published articles, mean (SD)	38.2 (39.7)	6.2 (10.0)	8.5 (12.2)	24.6 (33.9)	.42[c]
Familiarity with article submission,[e] mean (SD)	4.6 (0.8)	3.7 (0.7)	3.4 (1.3)	4.1 (1.1)	<.001[b]
Familiarity with peer review,[e] mean (SD)	4.2 (1.1)	2.8 (1.2)	2.5 (1.2)	3.5 (1.4)	<.001[b]
Confidence as peer reviewer,[e] mean (SD)	4.1 (1.1)	2.4 (1.2)	2.8 (0.9)	3.5 (1.3)	<.001[b]
January 2025 survey respondents, No.	20	11	12	43	.48[a]
Peer reviews issued, mean, (SD)	13.3 (37.4)	4.0 (5.8)	1.3 (1.4)	7.6 (25.9)	.009[c]
Confidence as peer reviewer,[e] mean (SD)	4.2 (0.6)	4.2 (0.6)	3.6 (1.0)	4.0 (0.8)	.063[b]
Pre- and post-event confidence as peer reviewer,[e] mean difference[d] (95% CI)	0.1 (–0.5 to 0.6)[f]	1.7 (0.9-2.6)[g]	0.8 (0.1-1.5)[h]	0.5 (0.1-1.0)[i]	NA
Checklist users, No. (%)	18 (90.0)	10 (90.9)	9 (75.0)	37 (86.1)	.52[a]
Checklist usefulness,[e] mean (SD)	4.5 (1.0)	4.9 (0.3)	4.4 (1.0)	4.6 (0.9)	.33[b]

Abbreviation: NA, not applicable.
[a]Fisher exact test.
[b]Analysis of variance.
[c]Kruskal-Wallis test.
[d]*t* test.
[e]5-Point Likert scale: 1, poor; 5, excellent.
[f]*P* = .80.
[g]*P* = .001.
[h]*P* = .03.
[i]*P* = .01.

reviews of full-length manuscripts. Reviews are graded by both mentors and assigned editors. Competitive program admission is by board member nomination and is limited to 45 individuals per year. From 2019 to 2025, 182 graduates completed the program. *Reviewer's Guild* members are primarily (83%) from the United States, but 10 other countries are represented. Surveys of each year's class are used for program improvement. Based on feedback, improvements included increased opportunities to learn statistical skills, networking, and leadership roles. The objective of this study was to report a separate 5-year survey to assess overall participant satisfaction with the program and career impact, particularly with respect to publishing, networking, and career opportunities.

Design This cohort study sought to evaluate the impact of a mentored peer review program. The STROBE[1] reporting guideline was followed. An anonymous online questionnaire regarding the program was developed and emailed to graduates. The summative survey evaluated nomination pathway, mentor participation, leadership, and networking opportunities resulting from program participation as well as publication output.

Results Ninety-six of the 182 graduates (52.8%) responded. Key survey responses are shown in **Table 25-0960**. Nearly all respondents reported enhanced understanding of the editorial process. Eighty-eight respondents (91.7%) had a generally positive experience and believed that the program improved the quality of their own publications and interpretation of others' work. Seventeen respondents (17.7%) said they had improved understanding of statistical principles. Thirty-seven respondents (38.5%) published 5 or more manuscripts since graduating from the *Reviewer's Guild*, 64 (66.7%) became involved with other journals (as ad hoc reviewers or editorial board members), and 16 (16.7%) of the highest performers received *Reviewer's Guild* board or full editorial board appointments to the journal.

Conclusions The *Reviewer's Guild* learning model appears to improve early career faculty and trainee peer review skills and leadership opportunities. While there are no comparisons with individuals who did not do the training, mentors saw improvement in review quality over the course of the program. In addition, graduates become loyal reviewers for the journal following graduation and are tracked as program graduates. As participants were surgeons, results may not be generalizable to other specialties. Such programs allow participants to become more involved in the editorial process and may potentially lead to journal leadership and career opportunities.[2,3]

References

1. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology

Table 25-0960. *Reviewer's Guild* Key Survey Results (N = 96)

Question	Response	No. (%)
Has your interpretation of manuscripts improved?	Agree/strongly agree	85 (88.5)
What skills do you feel that you most improved upon by participating in *Reviewer's Guild?*	Networking	39 (40.6)
	Understanding statistical procedures	17 (17.7)
	Leadership	19 (19.8)
	Editing publications (your own plus others)	83 (86.5)
	Writing papers for publication	71 (74.0)
Do you feel that *Reviewer's Guild* participation improved your own publications?	Agree/strongly agree	66 (68.8)
How many manuscripts have been published since graduation from *Reviewer's Guild* (print as well as e-pub)?	5-10	21 (21.9)
	10-15	11 (11.6)
	≥24	5 (5.2)
Have you become involved in another journal's peer review process, either via ad hoc reviewing or the editorial board since your time with *Reviewer's Guild?*	Yes	64 (66.7)
Did you become a DCR *Reviewer's Guild Associate* Editor?	Yes	12 (12.5)
Did you become an Associate Editor?	Yes	4 (4.2)
As a result of your time in *Reviewer's Guild,* have you been asked to perform any of the following?	Review abstracts for society program committee	29 (30.2)
	Serve as a moderator for annual society meeting	26 (27.1)
Has your participation in other leadership positions increased since your time in *Reviewer's Guild?*	Yes	35 (36.5)
Was your overall experience with *Reviewer's Guild* a positive one?	Yes	92 (95.8)
Did your mentor perform as expected during the *Reviewer's Guild* process?	Agree/strongly agree	80 (83.3)

DCR indicates *Diseases of the Colon & Rectum.*

(STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335(7624):806-808. doi:10.1136/bmj.39335.541782.AD

2. Isaacson RA, Bay SN, McCarty MM. Supporting the next generation of researchers: GENETICS peer review training program. *Sci Ed*. 2020;43:77. doi:10.36591/SE-D-4303-77

3. Lyons-Warren AM, Aamodt WW, Strowd R, Pieper KM, Merino JG. Assessment of a structured and mentored peer review curriculum on quality of peer review. Abstract presented at: Ninth International Congress on Peer Review and Scientific Publication; September 8-10, 2022; Chicago, IL. Accessed July 1, 2025. https://peerreviewcongress.org/wp-content/uploads/2022/11/Program2022.pdf

[1]Division of Colon & Rectal Surgery, University of Louisville, Louisville, KY, US, sogala01@louisville.edu; [2]Price Institute for Surgical Research, University of Louisville, Louisville, KY, US.

Conflict of Interest Disclosure Susan Galandiuk receives a stipend for services as editor of the journal *Diseases of the Colon & Rectum* from the American Society of Colon and Rectal Surgeons.

Errors and Corrections

In-person

Quotation Inaccuracy in Medicine: A Systematic Review and Meta-Analysis

Christopher Baethge,[1,2] Hannah Jergas[3]

Objective Although quotations are central to scientific publications, they often misrepresent the cited reference.[1] Inaccurate quotations may mislead readers, undermine the line of argument of a paper, and threaten the integrity of the scientific record.[2,3] We updated an earlier meta-analysis[1] to assess improvement in recent years.

Design This systematic review and meta-analysis, preregistered on Open Science Framework (OSF), was conducted following the Cochrane Handbook and PRISMA statement. It focuses on quotation inaccuracy—errors of content—as opposed to bibliographical errors, often referred to as citation errors. We searched MEDLINE, PubMed Central via PubMed, and Web of Science from January 1, 2014, through October 1, 2023, for studies on quotation errors in the medical literature, with no date, language, or source restrictions, and we hand-searched all reference lists of included studies. Literature screening, data extraction, and risk of bias estimation were carried out by 2 authors independently. We adopted the original studies' definitions[1] of major inaccuracies (ie, not at all in accordance with the claim of the original authors), minor inaccuracies (ie, inconsistencies and factual errors not severe enough to contradict a statement by citing authors), secondary quotations (ie, quotations of sources referencing an original study), and total quotation errors. Summary proportions of quotation errors were estimated in random-effects meta-analyses (log-transformed proportions and DL τ^2 estimates), with heterogeneity measures, 95% confidence intervals, and 95% prediction intervals. Time trends were investigated in meta-regressions.

Results Based on 46 studies included, with 32,074 quotations/references checked, 16.9% were incorrect (95% CI, 14.1%-20.0%). Of these, 8.0% (95% CI, 6.4%-10.0%) were major inaccuracies and 7.8% (95% CI, 5.7%-10.5%) were minor inaccuracies. Heterogeneity throughout all meta-analyses was high (**Table 24-0811**). Rates of total errors have not improved during the last decade (slope: −0.002; 95% CI, −0.03 to 0.02; $P = .85$) and neither have rates of major errors (slope: 0.002; 95% CI, −0.02 to 0.03; $P = .89$). There were no associations with risk of bias, publication bias, number of references, or medical specialty (surgical, nonsurgical, other), but the total error rate was statistically significantly ($P = .04$; binomial, n = 25) and negatively correlated with Journal Impact Factor (−0.25; Spearman, n = 23). Interrater reliability was sufficient (Cohen κ: 0.73), and summary estimates were supported by sensitivity analyses employing arcsine square root data transformation and Sidik and Jonkman τ^2 estimates.

Table 24-0811. Quotation Inaccuracy in Medicine: Main Results and Heterogeneity Estimates in Percentages (46 Studies With 32,074 Quotations/References Investigated)

Quotation error category (No. of studies)	%				
	Summary estimate	95% Confidence interval	95% Prediction interval	I^2	τ^2
Total errors (46)	16.9	14.1-20.0	5-46	98	0.493
Major errors (36)	8.0	6.4-10.0	2-26	94	0.461
Minor errors (36)	7.8	5.7-10.5	1-39	97	0.911
Secondary errors (18)	5.3	3.3-8.5	1-36	97	1.109

Conclusions Quotation inaccuracy remains prevalent in scientific medical texts. Despite seemingly broad awareness among journal editors and probably among many authors, there is no evidence of improvement in recent years. It is uncertain how the utilization of artificial intelligence in writing articles and in double-checking quotation inaccuracy will impact the problem, but human effort will likely remain important in creating and vetting quotes—on the part of authors, editors, reviewers, and readers.

References

1. Jergas H, Baethge C. Quotation accuracy in medical journal articles—a systematic review and meta-analysis. *PeerJ*. Published online October 27, 2015. doi:10.7717/peerj.1364

2. Baethge C. Importance, errors, and patterns of quotations to psychiatric original articles. *Pharmacopsychiatry*. 2020;53(6):247-255. doi:10.1055/a-1167-35673

3. Peoples N, Østbye T, Yan LL. Burden of proof: combating inaccurate citation in biomedical literature. *BMJ*. Published online November 6, 2023. doi:10.1136/bmj-2023-076441

[1]*Deutsches Ärzteblatt* and *Deutsches Ärzteblatt International*, Editorial Offices, Cologne, Germany, baethge@aerzteblatt.de; [2]Department of Psychiatry and Psychotherapy, Faculty of Medicine, University of Cologne, Cologne, Germany; [3]Department of Neurology, Faculty of Medicine, University of Cologne, Cologne, Germany.

Conflict of Interest Disclosures Both authors have published on this topic before. Christopher Baethge is employed by a general medical journal (*Deutsches Arzteblatt*, *Deutsches Arzteblatt International*).

Acknowledgment The authors gratefully acknowledge Joan Albert Hammerstein's help in screening the literature.

Additional Information This study is registered on Open Science Framework (OSF) at https://osf.io/95rej/?view_only=8c5d2b51a8814278bff1fec1e40d0bb9

Virtual

Taiwanese Researchers' Perceptions of Errors and Their Coping Strategies

Chien Chou[1]

Objective This study surveyed how frequently Taiwanese researchers identify errors in others' journal articles and their own, as well as how they address these errors.

Design This cross-sectional study used a self-designed questionnaire from June to December 2023, categorizing errors into 7 dimensions: research design (6 items), data and information (8 items), writing and citation (12 items), authorship (2 items), conflicts of interest (2 items), journal publication (3 items), and others (1 item), totaling 34 items. A 4-point Likert scale (often, occasionally, seldom, and never, ranging from 4 to 1) was used. The error-addressing methods were presented as questions, allowing multiple answers. Using snowball sampling, this study collected 593 valid responses from Taiwanese professors and researchers (312; 52.61%) as well as from PhD students with publication experience (281; 47.39%). Ethics approval was obtained from the National Yang Ming Chiao Tung University institutional review board before data collection. Content validity of the survey was confirmed by experts.

Results The participants reported a higher frequency of identifying errors in others' articles than those in their own (t = 17.88; 95% CI, 0.35-0.43; Cohen d = 0.67; P < .001). Regarding identifying others' errors, 22 items had a mean frequency above the theoretical mean of 2.5, in the categories of occasionally and seldom. The item with the highest frequency was spelling errors/omissions, followed by errors in the understanding or application of theory. For identifying one's own errors, only the item spelling errors/omissions reached the theoretical mean of 2.5. The other 33 items, such as misunderstanding of previous research and errors in understanding theory application, ranged from 2.34 to 1.92, corresponding to seldom and never. In terms of addressing errors (multiple responses allowed; see **Table 25-0868**), 478 participants (80.6%) indicated that if they identify others' errors, they would use them "as lessons to avoid making the same mistakes in the future," 315 (53.1%) would take the errors "as examples to warn others," and 179 (30.2%) would "just observe the errors and take no action." As for addressing their own errors, 253 participants (42.7%) would "submit a proposed correction notice to the journal," 232 (39.1%) would "take no action to correct the errors and use them as lessons," 166 (28.0%) would take no action and use the errors "as examples to warn others."

Conclusions Taiwanese researchers tend to identify errors more frequently in others' work, viewing their own errors as cautionary examples for themselves rather than warnings for others. While correcting errors is a positive step, the tendency to be observant of others' errors and more reflective to correct their own suggest that Taiwanese researchers tend to take errors more personally. The finding highlights the Committee on Publication Ethics principles of transparency and underscores the significance of upholding research integrity.

[1]Institute of Education, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, cchou@nycu.edu.tw.

Conflict of Interest Disclosures None reported.

Table 25-0868. Methods for Addressing Others' and Their Own Errors (N = 593, Multiple Responses Allowed)

If identifying errors in others' journal articles, I would…		If identifying errors in my own journal articles, I would…	
Item	**No. (%)**	**Item**	**No. (%)**
Just observe the errors and take no action	179 (30.2)	Just observe the errors and take no action	82 (13.8)
Use the errors as lessons to avoid making the same mistakes in the future	478 (80.6)	Take no action to correct the errors and use them as lessons to avoid making the same mistakes in the future	232 (39.1)
Take the errors as examples to warn others (such as students) not to make the same mistakes	315 (53.1)	Take no action to correct the errors and use them as examples to warn others (such as students) not to make the same mistakes	166 (28.0)
Address them in my own subsequent research paper and highlight the flaws	132 (22.3)	Take no action to correct the errors but address them in my own subsequent research paper and highlight the flaws	113 (19.1)
Inquire of the author regarding the errors	67 (11.3)	Take no action but request a correction from the journal if others spot the errors	61 (10.3)
Report the article with errors to the journal	30 (5.1)	Submit a proposed correction notice to the journal	253 (42.7)
Report the article with errors to relevant institutions as an allegation of research integrity	20 (3.4)	Submit a proposed retraction notice to the journal	95 (16.0)
Discuss the errors on academic social media platforms or peer review sites such as PubPeer	79 (13.3)	NA	NA
Take other actions instead of the ones above	4 (0.7)	Take other actions instead of the ones above	27 (4.6)

Abbreviation: NA, not applicable.

Funding/Support This work was supported by the National Science and Technology Council, Taiwan (MOST110-2511-H-A49-008-MY4 and NSTC113-2750-V-A49-001-MY2).

Role of the Funder/Sponsor The funder provided financial support only and had no role in the preparation of the abstract proposal.

Funding/Grant Peer Review

In-person

Multistakeholder Perspectives on Current Attitudes Toward (Un)masking Reviewers' Identity in Biomedical Research Proposals' Peer Review: A Qualitative Study

Seba Qussini,[1,2] Farizah Mezer Anami,[2] Kris Dierickx[1]

Objective Many peer review attributes are widely criticized and remain poorly investigated, particularly in the context of proposals' peer review.[1] This study aims to explore the diverse perspectives of stakeholders regarding the role of unmasking in funding proposals' peer review and the implications of open peer review as part of the recent open science movement, specifically in biomedical research proposals' peer review.

Design To describe participants' perspectives as constructed through their recent experiences, we have conducted a generic descriptive qualitative study[2] within a constructivist paradigm, using semistructured interviews to gather insights from reviewers, applicants, and peer review scholars. A total of 23 participants were selected through purposive and snowball sampling from funding agencies in Belgium and Qatar with whom no prior relationship had been established. Interviews were conducted between June 2024 and February 2025 by the first author (S.Q.). Transcribed interviews were analyzed according to the 6-step thematic framework analysis described by Braun and Clarke.[3] Initially, autogenerated transcripts were read and checked for in-depth familiarization with the data, which was followed by a line-by-line inductive coding, conducted iteratively after each set of 2 interviews. We followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines.

Results Codes with shared characteristics were grouped into categories, and ultimately 3 overarching themes were generated: (1) the importance of increased transparency in fund allocation procedures while maintaining anonymized reviewer identities; (2) open peer review as a feasible approach for enhancing transparency and accountability in funding proposals' peer review; and (3) a growing critical stance toward traditional peer review systems, calling for alternative models like baseline or lottery funding procedures. Collectively, the results shed light on the perceived advantages and limitations of different peer review models and provide an understanding of how unmasking identities influences the fairness and objectivity of fund allocation decisions.

Conclusions There is persistent preference for double-anonymous review among the scientific community; however, researchers are increasingly aware of the shortcoming of anonymized review, especially in light of current challenges within the funding landscape. Simultaneously, they recognize the importance of greater openness in peer review and increased transparency in fund allocation procedures.

References

1. Qussini S, MacDonald RS, Shahbal S, Dierickx K. Blinding models for scientific peer-review of biomedical research proposals: a systematic review. *J Empir Res Hum Res Ethics*. 2023;18(4):250-262. doi:10.1177/15562646231191424

2. Qussini S, Dierickx K. Multi-stakeholder perspectives on current attitudes toward (un)blinding reviewers' identity in biomedical research proposals peer review: a qualitative study. OSF Registries. https://osf.io/dp65f

3. Clarke V, Braun V. Thematic analysis. *J Pos Psychol*. 2017;12(3):297-298. doi:10.1080/17439760.2016.1262613

[1]Centre for Biomedical Ethics and Law, Faculty of Medicine, KU Leuven, Leuven, Belgium, seba.qussini@student.kuleuven.be; [2]Hamad Medical Corporation, The Medical Research Center, Doha, Qatar.

Conflict of Interest Disclosures None reported.

Funding/Support The publication and conference fees will be funded by the Medical Research Center at Hamad Medical Corporation through grant MRC-01-24-305.

Role of the Funder/Sponsor The Medical Research Center had no role in the design, analysis, or interpretation of this study.

Influence of Using a Systematic Review to Justify New Research on Funding Application Score

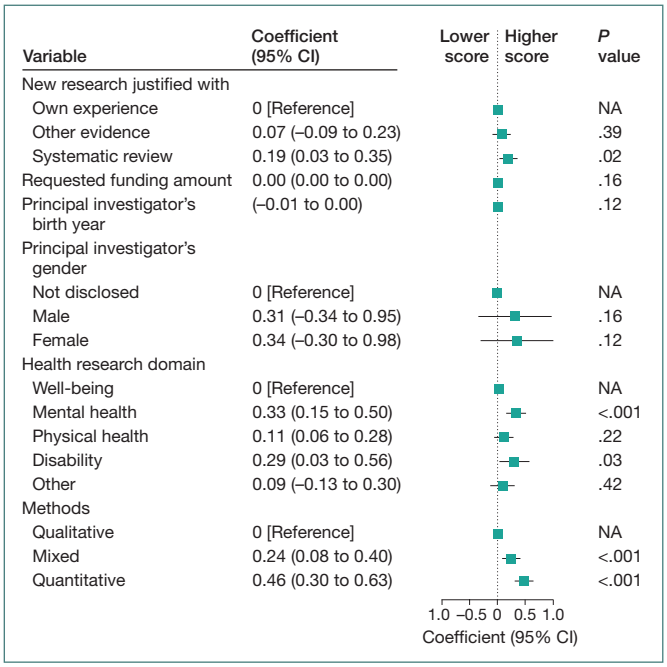
Jong-Wook Ban,[1] Hans Lund,[2] Karen Robinson,[3] Ida Svege,[4] Jan-Ole Hesselberg[5,6]

Objective Many have argued that the need for new research should be determined through a systematic examination of existing evidence to maximize its value.[1,2] Our study evaluated the outcomes of following such arguments on funding applications. We tested a hypothesis that funding applications justifying new research with a systematic review would receive higher scores than those without such justification.

Design We analyzed funding applications submitted to Stiftelsen Dam, one of Norway's largest health foundations, during the 2024 cycle. We used multiple regression to test our hypothesis while controlling for the influence of the following 5 covariables: requested funding amount, principal investigator's birth year, principal investigator's gender, project's health domain, and research methods.

Results Of 420 applications received, 193 (46.0%) cited a systematic review, 181 (43.1%) referenced a published or unpublished study, and 46 (11.0%) referred to applicants' own experience as justification for new research. The median requested funding amount was kr3,000,000 ($267,796), with an IQR of kr2,801,000 to kr3,000,000. There were 158 (37.6%) men and 247 (58.8%) women principal investigators, and 15 (3.6%) who did not disclose their gender. The median birth year of the principal investigators was 1976 (IQR, 1968-1982). Funding applications included research projects in mental health (122 [29.0%]), physical health (197 [46.9%]), well-being (42 [10.0%]), disability (18 [4.3%]), and other domains (41 [9.8%]). Qualitative, mixed, and quantitative methods were used in 53 (12.6%), 145 (34.5%), and 222 (52.9%) applications, respectively. The median funding application score was 4.8 (IQR, 4.4-5.1) on a 7-point scale, where 1 represents poor and 7 represents excellent. Applications that cited a systematic review as justification received scores 0.19 (95% CI, 0.03-0.35) points higher than those that relied on the applicants' own experience (**Figure 25-1129**). This could have raised an application with an average score by approximately 100 places in a ranking of 1716 applications. Applications that cited a systematic review received scores 0.12 (95% CI, 0.02-0.22) points higher than those that cited other types of evidence.

Figure 25-1129. Association of Variables With Review Score



NA indicates not applicable.

Conclusions Our exploratory analysis showed an association between citing a systematic review and higher funding application scores. Our study has the following limitations. First, we could only control for variables available in our data. So, it was impossible to exclude the influence of other potential confounders, such as study design. Second, we evaluated funding application scores rather than funding outcomes. We could not evaluate the funding outcomes due to the very small number of applications receiving funding each year. Because Stiftelsen Dam heavily relies on funding application scores to decide on funding, we used those scores as a surrogate for funding outcomes.

References

1. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet*. 2014;383(9912):156-165. doi:10.1016/S0140-6736(13)62229-1

2. Lund H, Brunnhuber K, Juhl C, et al. Towards evidence based research. *BMJ*. 2016;355:i5440. doi:10.1136/bmj.i5440

[1]Section of Hospital Medicine, University of Chicago, Chicago, IL, US, jban@uchicago.edu; [2]Section of Evidence-Based Practice, Western Norway University of Applied Sciences, Bergen, Norway; [3]Johns Hopkins Evidence-Based Practice Center, Johns Hopkins University, Baltimore, MD, US; [4]Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway; [5]Stiftelsen Dam, Oslo, Norway; [6]Department of Psychology, University of Oslo, Oslo, Norway.

Conflict of Interest Disclosures Ida Svege reported being the former Head of Program Development (Leder for programutvikling) of Stiftelsen Dam. Jan-Ole Hesselberg reported being the current Program Manager (Programsjef) of Stiftelsen Dam.

Virtual

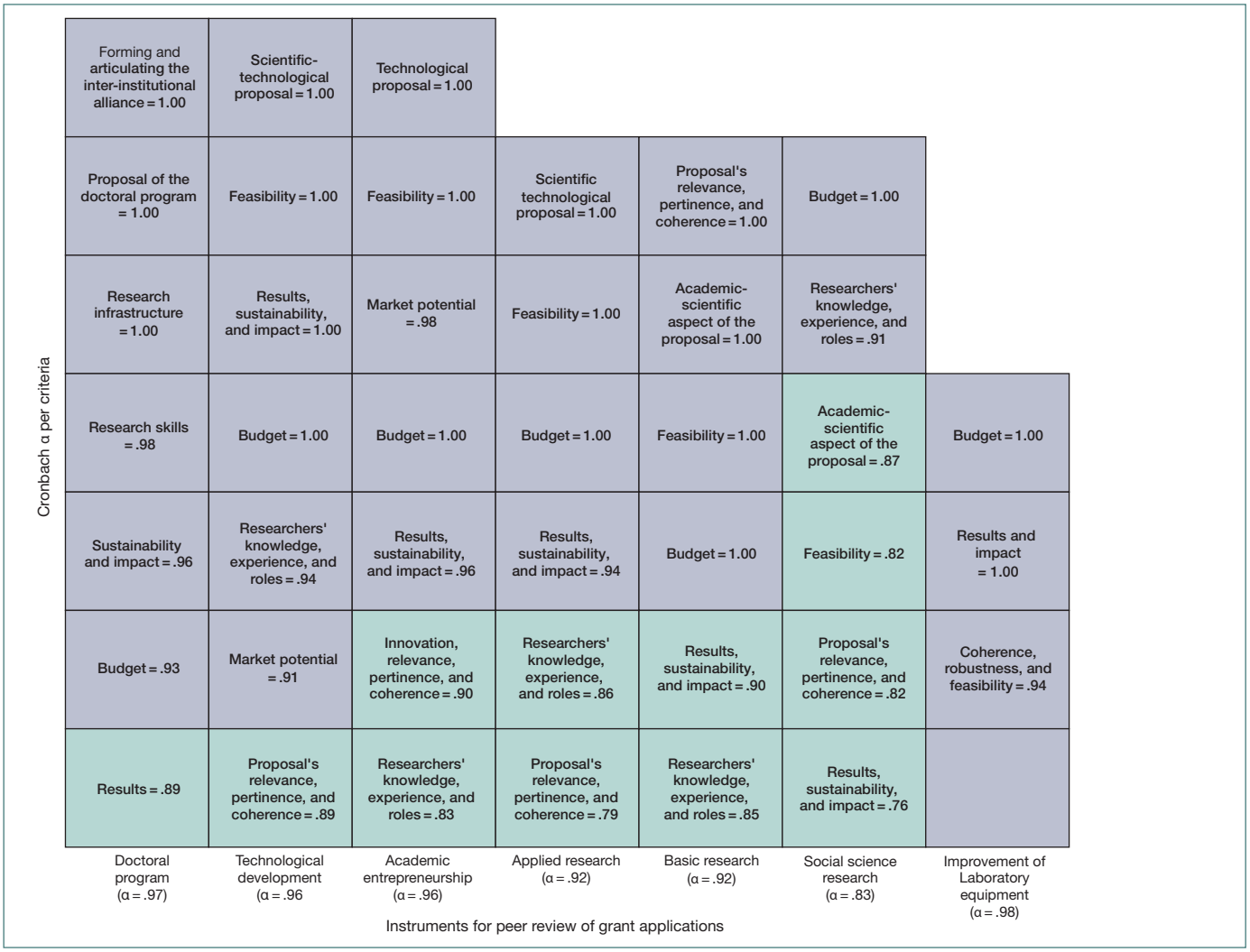
Construction and Validation of Instruments for the Peer Review of Grant Proposals in Peru

Max Carlos Ramírez-Soto,[1] Laura Alvarado-Barbarán,[1] Dianeth Rojas-Naccha,[1] Ayda Luna-Mercado,[1] Arlet Arce-Zavala[1]

Objective Peer review plays a central role in selecting research proposals to obtain grants from science, technology, and innovation (STI) agencies.[1] Different criteria and questions are used during this process.[2] Given the time and resources involved, there is no available information on the reliability of the tools and review criteria used in the peer review of grant proposals. Peru's National Program for Scientific Research and Advanced Studies (PROCIENCIA) uses a 2-tiered review process: an initial peer review to assess scientific merit, followed by a panel review to evaluate proposal relevance in line with funding priorities. Expert scientists are involved in both stages of the review process. Proposals may be improved prior to the final allocation of funding. In that context, we assess the reliability of the instruments used for the peer review of grant proposals at PROCIENCIA.

Design This cross-sectional study assessed the construction and validation of 7 instruments used by PROCIENCIA for the peer review of grant applications between January 2022 and December 2024. Peer review instruments were applied in 7 funding calls (**Figure 25-1021**). Each financial instrument has distinct requirements and aims. The process of constructing and validating the instruments was carried out in the following stages: (1) literature review, (2) instrument construction, and (3) content validation. The literature search included studies on the criteria for reviewing STI grant applications, including the review criteria used by STI agencies. The findings from the literature review were used as evidence for the development of the evaluation criteria and questions included in the peer review instruments. We used a modified Delphi process to validate the content,[3] which included a review round by 8 to 15 expert panelists for each instrument. The experts were senior researchers with extensive experience in reviewing grant applications and scientific articles in various fields of scientific and technological knowledge. Each instrument contained 4 to 7

Figure 25-1021. Reliability Index of the Instruments Used by PROCIENCIA for the Peer Review of Grant Proposals



LMICs indicates low- and middle-income countries.

criteria (eg, knowledge of the researchers, relevance and coherence of the proposal, the scientific-technological aspect of the proposal, feasibility, results, and budget) as well as 3 to 8 questions. Content validation was performed at the overall instrument level (overall score) and for each criterion (individual score). Cronbach α was used to assess internal consistency; a value greater than .75 was considered acceptable.

Results The 7 instruments used for the peer review of grant applications had an overall reliability index between .86 and .98 (**Figure 25-1021**). All instruments had acceptable reliability indices regarding assessment criteria (>.75). The assessment criteria of the doctoral program, technological development, and improvement of laboratory equipment instruments had a higher reliability index (range, .89-1.00 and .92-1.00, respectively).

Conclusions Our results show that PROCIENCIA uses validated and reliable instruments for the peer review of grant proposals. Validation and reliability of peer review instruments for STI funding is helpful in ensuring the integrity of the review process and in selecting research proposals for funding. Therefore, funding agencies should consider validating peer review instruments.

References

1. Hug SE, Aeschbach M. Criteria for assessing grant applications: a systematic review. *Palgrave Commun*. 2020;6(1):1-15. doi:10.1057/s41599-020-0412-9

2. Abdoul H, Perrey C, Amiel P, et al. Peer review of grant applications: criteria used and qualitative study of reviewer practices. *PLoS One*. 2012;7(9):e46054. doi:10.1371/journal.pone.0046054

3. Hsu C, Sandford BA. The Delphi technique: making sense of consensus. *Practl Assess Res Eval*. 2007;12(1):10. doi:10.7275/pdz9-th90

[1]Beneficiary Selection Sub-Unit, National Program of Scientific Research and Advanced Studies (PROCIENCIA), National Council for Science, Technology and Innovation (CONCYTEC), Lima, Peru, mramirez@prociencia.gob.pe.

Conflict of Interest Disclosures None reported.

Acknowledgment The authors would like to thank Sixto Sánchez, president of CONCYTEC, and Dora Blitchtein Winicki De Levy, executive director of PROCIENCIA, for the technical support for the study.

Reviewers' Interpretation and Application of Research Quality Criteria in Grant Peer Review

Rachel Claus[1]

Objective Research quality criteria guide grant applications and reviewer evaluations. When research crosses disciplinary boundaries, it requires more expansive quality criteria. Research evaluation is challenging in this context because concepts of research quality are rooted in disciplinary tradition.[1] Reviews are characterized by inconsistency and low agreement[2] and exacerbated by poorly defined criteria.[3] This presentation focuses on how consistently quality criteria are applied and interpreted by reviewers and explores reasons for inconsistency.

Design All scoring data were collected from 2 noncompetitive review processes of multimillion-dollar research program proposals with aims to improve food security. In 2021, 32 research proposals were evaluated using 17 criteria on a standard 4-point Likert scale (0-3), with limited reviewer overlap. In 2024, 9 proposals were evaluated using 12 criteria, and 4 using 11 criteria, with no reviewer overlap. Each proposal was assessed by a panel of 3 reviewers who scored independently before discussing scores to reach consensus. In total, 45 proposals were evaluated by 80 individual subject matter experts across 45 panels. Score consistency was measured by discrepancy per criterion as the difference between the highest and lowest score awarded in a panel, and how many reviewers agreed on their individual assessments. The standard of consistency was met when at least 2 of 3 reviewers agreed, and the discrepancy between individual reviewer scores was less than or equal to 1 Likert scale point. A total of 696 panel-level measures of consistency were computed. Cross tabulations were used to identify frequencies of inconsistency by criterion. Interviews with reviewers were conducted to understand perceived reasons for individual review discrepancies and disagreement, and individual and panel-level score justifications were analyzed to explore criteria interpretations.

Results There was a statistically significant relationship between consistency and the evaluation criteria ($\chi^2 = 61.013$; $P < .001$). Some criteria were more consistently applied than others. The frequency that each criterion met the standard of consistency is presented in **Table 25-1061**. Reviewers more frequently applied the following criteria inconsistently when evaluating proposals: comparative advantage (26.7%), monitoring, evaluation, and learning (24.4%), and overall theory of change (24.4%). Justified and transparent costing had the highest rate (71.9%) of inconsistency in 2021 but was not evaluated in the 2024 review cycle. Individual score discrepancies and disagreement were perceived by reviewers to result from diverse disciplinary expertise and a constructive way to achieve comprehensive quality assessments. More problematic reasons for discrepancies included misalignment of criteria to the application and different interpretations resulting from individual values and perceived abilities to make judgments.

Conclusions The preliminary results indicate scope for criteria clarification to improve consistency in interpretation and reliability of individual reviewer's quality assessments of proposals. The approach can be adapted to test and inform improvements to quality criteria.

References

1. Defila R, Di Giulio A. Transdisciplinary development of quality criteria for transdisciplinary research. In: Regeer BJ, Klaassen P, Broerse JEW, eds. *Transdisciplinarity for*

Table 25-1061. Cross Tabulation of Criteria and Consistency Standard Met (Yes/No)[a]

Criteria	Consistency standard, count (%)		Total No.
	Panels that scored criterion inconsistently	Panels that scored criterion consistently	
Research problem	8 (17.8)	37 (82.2)	45
Demand driven	6 (13.3)	39 (86.7)	45
Research alignment	8 (17.8)	37 (82.2)	45
Overall theory of change	11 (24.4)	34 (75.6)	45
Work package theory of changes	2 (4.4)	43 (95.6)	45
Research methods	8 (17.8)	37 (82.2)	45
Tradeoffs and synergies	12 (37.5)	20 (62.5)	32
Impact at scale	13 (28.9)	32 (71.1)	45
Gender and social inclusion	10 (24.4)	31 (75.6)	41
Risk framework	10 (22.2)	35 (77.8)	45
Comparative advantage	12 (26.7)	33 (73.3)	45
Capacity building	9 (28.1)	23 (71.9)	32
Project management	7 (21.9)	25 (78.1)	32
Justified and transparent costing	23 (71.9)	9 (28.1)	32
Outputs and public sharing	8 (17.8)	37 (82.2)	45
Monitoring, evaluation, and learning	11 (24.4)	34 (75.6)	45
Evaluation and impact assessment	9 (28.1)	23 (71.9)	32
Total	167 (24.0)	529 (76.0)	696

[a]The criteria for proposal assessment have been adapted from the Independent Science for Development Council from April 2021.

Transformation. Palgrave Macmillan, Cham; 2024. https://doi.org/10.1007/978-3-031-60974-9_5

2. Pier EL, Brauer M, Filut A, et al. Low agreement among reviewers evaluating the same NIH grant applications. *Proc Natl Acad Sci U S A*. 2018;115(12):2952-2957. doi:10.1073/pnas.1714379115

3. Abdoul H, Perrey C, Amiel P, et al. Peer review of grant applications: criteria used and qualitative study of reviewer practices. *PLoS One*. 2012;7(9):e46054. doi:10.1371/journal.pone.0046054

[1]Royal Roads University, Victoria, BC, Canada, rachel.claus@royalroads.ca.

Conflict of Interest Disclosures None reported.

Funding/Support This research was supported by the Sustainability Research Effectiveness Program, the BC Graduate Scholarship, the Royal Roads Doctoral Scholarship, and the David Harris Flaherty Scholarship.

Role of the Funder/Sponsor The research was done as part of the Sustainability Research Effectiveness Program, with input provided to the design of the study, review and approval of the abstract and input to the decision to submit the abstract for presentation. Scholarships provided general support without any role in design and conduct of the study; collection, management and interpretation of the data; preparation, review or approval of the abstract, and decision to submit the abstract for presentation.

Experiences and Challenges Faced by Canadian Health Research Grant Peer Reviewers

Joanie Sims Gould,[1] Anne Lasinsky,[2] Adrian Mota,[3] Karim M. Khan,[1,2,4] Clare L. Ardern[5,6]

Objective There is robust debate about the perceived strengths and weakness of grant peer review. Much of the research on issues in grant peer review is based on quantitative analysis of funding or scoring outcomes, which does not illuminate the experiences of peer review committee members. The objective of this study was to explore and understand the experiences and challenges faced by Canadian health research grant peer reviewers.

Design This qualitative study received ethics approval from the University Behavioural Research Ethics Board. Study conduct and reporting followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) guideline. Chairs, peer reviewers, and Scientific Officers of the Canadian Institutes of Health Research (CIHR) project grant competition peer review panels were interviewed. CIHR staff prepared a list of 50 potential participants who represented the 4 primary branches of CIHR research (biomedical, clinical, health systems and services, population health) from a public website. The inclusion criterion was having participated in a CIHR Project Grant competition peer review panel at least once as a peer reviewer, chair, or scientific officer. The researchers randomly selected names from the list and sent a recruitment email inviting participants to an online semistructured interview. The response rate was 36%. Two experienced qualitative researchers recruited and interviewed participants on a rolling basis from February to August 2022. The study team met biweekly to review the interview transcripts. All participants provided verbal consent for audio recording and reporting of quotes at the beginning of the interview. The interview guide was developed based on a priori concepts of peer review and the study team's research experience in the field of grant peer review. There were questions about participants' background, training in peer review, strengths and challenges of the review process, and managing conflict and bias. The analysis used a framework analysis approach.[1] Data were sifted, charted, and sorted based on key issues and themes to identify a thematic framework, compare and contrast themes, and explore similarities and differences.

Results Eighteen participants were interviewed, all of whom were mid- or senior-career researchers (age 42-77 years); 11 participants (61.1%) were women and 7 (38.9%) were men. Twelve participants (66.7%) identified as White, 3 (16.7%) as South Asian, and 3 (16.7%) as other race or ethnicity. Participants identified 3 threats to grant peer review: (1) lack of training and limited opportunities to learn, (2) challenges in differentiating and rating applications of similar strength, and (3) relying on academic reputations and personal

relationships in the review process to differentiate grant applications of a similar rating.

Conclusions Experienced grant peer reviewers in the Canadian health funding context identified an absence of training and learning opportunities for peer review, difficulty differentiating between applications of similar strength, and an emphasis on academic reputations and personal relationships in rating applications for the Project Grant Competition.

Reference

1. Srivastava A, Thomson SB. Framework analysis: a qualitative methodology for applied policy research. *Journal of Administration and Governance*. 2009;72. Accessed May 31, 2025. https://ssrn.com/abstract=2760705

[1]Department of Family Practice, The University of British Columbia, Vancouver, British Columbia, Canada; [2]School of Kinesiology, The University of British Columbia, Vancouver, British Columbia, Canada; [3]Canadian Institutes of Health, Ottawa, Ontario, Canada; [4]Canadian Institutes of Health–Institute of Musculoskeletal Health and Arthritis, Vancouver, British Columbia, Canada; [5]Department of Physical Therapy, The University of British Columbia, Vancouver, British Columbia, Canada, clare.ardern@ubc.ca; [6]Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Victoria, Australia.

Conflict of Interest Disclosures Adrian Mota is acting vice president, Research–Programs for the Canadian Institute of Health Research (CIHR). Karim M. Khan is scientific director for the CIHR Institute of Musculoskeletal Health and Arthritis (2017-2025). No other disclosures were reported.

Funding/Support This work was supported by a CIHR research operating grant (scientific directors) held by Karim M. Khan.

Role of the Funder/Sponsor The CIHR had no role in design and conduct of the study. The CIHR did not participate in interpretation of the data or the preparation of the abstract and did not participate in the decision to submit the abstract for presentation.

Misconduct and Research Integrity

In-person

Experience With 12 Years of Plagiarism and Duplication Screening

Markus K. Heinemann,[1] Andreas Boening,[2] Kazunori Okabe,[3] Jessica Bogensberger[4]

Objective In 2013, a cardiothoracic surgical journal introduced plagiarism detection (Crossref, powered by iThenticate). After a phase of 2 years during which all submissions were screened, since January 2015 only returned manuscripts after revision were checked by default. Primary submissions were then screened at the discretion of the Editor in Chief (EiC) only. The aim of this study was to report the 12-year experience.

Design All manuscripts submitted between January 1, 2013, and December 31, 2024, and branded by the plagiarism search engine with a similarity index above 33% were analyzed regarding the originator of the screening (EiC or revision default), reasons (actual plagiarism or self-plagiarism, which is a form of duplication), similarity percentage, and final decision. Descriptive statistics were applied. All updates of the search tool were implemented.

Results During the investigation period, 4083 manuscripts were submitted. A total of 99 (2.4%) showed a similarity index greater than 33%. The mean percentage was 53.3% (minimum, 34%; maximum, 99%; median, 50%; SD, ±14.8). Plagiarism and duplication search was prompted by the EiC in 48 papers and by automated default in 51 papers. Self-plagiarized or duplicated content was by the same authors in 71 of 99 papers (71.7%), by different authors in 18 (18.2%), and both in 10 (10.1%). Fifty-five manuscripts (55.6%) remained unpublished: 17 rejected without review, 34 rejected, and 4 withdrawn. Forty-four of 99 (44.4%) were accepted after revisions. There were 4 attempts at duplicate publication, all rejected. Preprints were discovered in 3 cases, with 1 accepted after revision: the preprint had been uploaded by a different journal for review and was rejected there, but then not deleted. The 44 manuscripts finally accepted contained previously published sequences by the same authors in 36, by different authors in 4, and both in 4. In 25 cases, the EiC rendered the similarity uncritical (eg, materials and methods section). The most frequent necessary correction was adequate citation. No adequate citation was given for self-plagiarized/duplicated content in 23 cases and for content by different authors in 4.

Conclusions Although the overall rate of plagiarism and duplication can be considered low, scientific misconduct was detected and its publication prevented in a few cases. In many instances, duplicated passages contained similar material and methods previously published by the same group. Deficiencies in adequate citation are a problem and necessitate continued education of authors. Selective plagiarism and duplication screening remains an important tool, although chosen thresholds of similarity remain arbitrary. Further improvements of sensitivity are expected, especially with the implementation of machine learning advances. Editorial decisions, however, cannot be made on automated findings alone, but must be put into personal perspective.

[1]German Society for Thoracic and Cardiovascular Surgery (DGTHG), Germany, heinemann@uni-mainz.de; [2]Universitaetsklinik Giessen, Germany; [3]Bell Land General Hospital, Osaka, Japan; [4]Thieme Publishers, Stuttgart, Germany.

Conflict of Interest Disclosures Jessica Bogensberger is an employee of Thieme Publishers, the publishing house of the journal investigated. All other authors have no conflict of interest to declare.

Vulnerability of Automated Text Matching–Based Reviewer Assignments to Collusions

Jhih-Yi Hsieh,[1] Aditi Raghunathan,[1] Nihar B. Shah[1]

Objective Collusion rings—researchers who manipulate reviewer assignments to review each other's work—pose major challenges for machine learning and artificial intelligence conferences. As submissions grow substantially,

automated reviewer assignment algorithms[1] have become common, often relying on the text similarity of reviewers' past papers and the author's submission.[2] While this text similarity is generally considered collusion safe,[3] this work investigated potential manipulations and proposes mitigations.

Design For a colluding author and reviewer, we designed a realistic 2-step attack procedure: (1) reviewer profile curation, in which the colluding reviewer kept only 1 past paper that was the most similar to the author's submission for text matching; and (2) submission abstract modification, in which the colluding author modified their submission's abstract by adding background sentences related to the themes in the reviewer's profile and inserting keywords that increased the text similarity to the reviewer. For scalability, we used a large language model (gpt-4-0125-preview [OpenAI]) to modify abstracts, although colluding authors in practice could manually refine them to reduce suspicion. We curated a dataset (3123 papers; 7900 reviewers) of the NeurIPS 2023 conference, which reviewed full papers (not abstracts). We then sampled reviewer-paper pairs where the reviewer initially ranked 20th, 101st, 501st, or 1001st among all reviewers by similarity, simulated collusion via our attack, and reported if the reviewer became highly ranked afterward. We proposed 2 hypotheses to improve robustness. (1) Requiring more past papers in reviewer profiles reduces attack success. (2) When a reviewer has multiple past papers in their profile, the similarities between each past paper and the submission are first calculated, then aggregated by taking the mean or maximum. We hypothesized that mean aggregation is more robust than maximum because averaging reduces manipulation effects. We tested these hypotheses via 2 ablations: (1) enforcing 10 papers in reviewer profiles and (2) comparing mean vs maximum under attack. Lastly, a randomized controlled trial tested for the identifiability of adversarial abstracts by unsuspecting human reviewers. Participants were randomly assigned either the benign or adversarial abstract to review.

Results The attack was highly effective (**Table 25-0911**); 67% of reviewers initially ranked 1001st became top 5, despite not working on similar topics. Both hypotheses on robustness were confirmed. When reviewers selected 10 past papers (100 samples), only 19% of similarity rankings were in the top 5 after the attack. For mean vs maximum aggregation (100 samples each), 32% and 49% of rankings were in the top 5 after attack, respectively. Lastly, in the randomized controlled trial (116 samples), while participants had more complaints for manipulated papers, surprisingly, benign papers elicited similar complaints, suggesting plausible deniability.

Conclusions This study found that conference reviewer assignments based on text matching are vulnerable to colluders and identified methods to improve robustness.

Table 25-0911. Attack Success Rates for Colluding Reviewers With Rankings Before Attack

Ranking before attack	Attack success rates ± SE (absolute No. of successful attacks)			Rankings after attack
	Top 1	Top 3	Top 5	Mean (95% CI)
20	90% ± 3% (90)	96% ± 2% (96)	98% ± 1% (98)	1.28 (1.06-1.50)
101	74% ± 3% (221)	89% ± 2% (267)	93% ± 2% (278)	2.22 (1.80-2.64)
501	60% ± 5% (60)	76% ± 4% (76)	83% ± 4% (83)	6.58 (3.47-9.69)
1001	48% ± 5% (48)	63% ± 5% (63)	67% ± 5% (67)	15.68 (6.82-24.54)

Abbreviation: SE, standard error.

The attack success rates are the fraction of times when the reviewers' rankings after attack were within the top 1000 among all reviewers ($k \in \{1,3,5\}$). The sample size is 100 for each row, except for when the ranking before attack is 101, where the sample size is expanded to 300, requested by a reviewer of this work.

References

1. Shah NB. Challenges, experiments, and computational solutions in peer review. *Commun ACM*. June 2022. Accessed July 11, 2025. https://www.cs.cmu.edu/~nihars/preprints/SurveyPeerReview.pdf

2. Cohan A, Feldman S, Beltagy I, Downey D, Weld D. SPECTER: document-level representation learning using citation-informed transformers. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020, 2270-2282. doi:10.18653/v1/2020.acl-main.20

3. Wu R, Guo C, Wu F, Kidambi R, Van Der Maaten L, Weinberger K. Making paper reviewing robust to bid manipulation attacks. In: *International Conference on Machine Learning*. 2021: 11240-11250.

[1]Carnegie Mellon University, Pittsburgh, PA, US, jhihyi.hsieh@gmail.com.

Conflict of Interest Disclosures Jhih-Yi Hsieh, Aditi Raghunathan, and Nihar B. Shah are employees of Carnegie Mellon University. Nihar B. Shah 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 J.P. Morgan Research Scholar Program (ONR N000142212181, NSF 2200410, 1942124, 2310758), the AI2050 program at Schmidt Sciences (G2264481), the Google Research Scholar Program, Apple, Cisco, and Open Philanthropy.

Role of the Funder/Sponsor The funders played no role in the design and execution of the experiment, data analysis, or preparation of the abstract.

Additional Information gpt-4-0125-preview (OpenAI) was used between April and September 2024 to generate modified abstracts without human supervision. Jhih-Yi Hsieh takes responsibility for the integrity of the content generated.

Identifying Potential Duplicate Publications in the Scientific Literature Using Crossref

Cyril Labbé,[1] Qinyue Liu,[1] Amira Barhoumi,[1] Olessya Miroshnichenko[1]

Objective For various reasons, duplicate publications are often considered problematic.[1,2] We tested a method to

semiautomatically detect such publications from publicly available metadata registered with Crossref by publisher.[3]

Design We first queried Crossref using ISSNs or ISBNs to build a set of (random) digital object identifiers (DOIs) with abstracts for duplicates to be searched. For each DOI, Crossref was queried again to retrieve a new set of 200 DOIs with similar titles containing potential duplicates of publications in the first set. We then identified abstracts in the second set that were nearly identical to those in the first set, differing by only a few words. When DOIs of apparent duplicates were not registered by preprint platforms (eg, bioRxiv, medRxiv), we would then automatically analyze when DOIs were registered by a different publisher, had differences in authorship, or were published in journals or books. The procedure was applied to the following sets of DOIs chosen to represent a wide variety of publishers for which abstracts are registered with Crossref: *Science* publications (2023-2024), *PLOS One* (2023-2024), *International Journal of Molecular Sciences* (2024), *BioMed Research International* (2024), *Scientific Reports* (2024), and IGI Global (DOIs included in books published in 2024).

Results The vast majority of duplicates found were not problematic, as they typically originated from preprint versions (**Table 25-0986**). For example, of 375 duplicated abstracts for articles published in *PLOS One*, 4 were not preprint publications. Moreover, close inspection revealed that 3 duplicates were republication of abstracts only. Only the 1 remaining abstract could be seen as problematic, as it was a duplication of the same exact content in 2 different journals, both open access.

Conclusions One limitation of the process is that despite having the exact same abstract, the body of the publications might be different. For many publications, abstracts are not registered with Crossref or sometimes are registered with errors. It is also difficult to identify when one of the DOIs represents a republication with an abstract alone. Abstracts publicly available at Crossref can be used to detect problematic duplicate publications as a cost-effective alternative to plagiarism detection tools. This experiment emphasizes the importance for publishers to answer positively to the Initiative for Open Abstracts (I4OA), which calls on all scholarly publishers to open the abstracts of their published works and, where possible, to submit them to Crossref.

References

1. Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ*. 1997;315(7109):635-640. doi:10.1136/bmj.315.7109.635

2. Errami M, Hicks JM, Fisher W, et al. Déjà vu—a study of duplicate citations in Medline. *Bioinformatics*. 2008;24(2):243-249. doi:10.1093/bioinformatics/btm574

3. van Eck NJ, Waltman L. Crossref as a source of open bibliographic metadata. *MetaArXiv*. Preprint posted online May 12, 2025. doi:10.31222/osf.io/smxe5

[1]Université Grenoble Alpes, French National Centre for Scientific Research, Grenoble INP, Laboratoire d'Informatique de Grenoble, Grenoble, France, cyril.labbe@imag.fr.

Conflict of Interest Disclosures None reported.

Funding/Support We acknowledge the NanoBubbles project that has received Synergy grant funding from the European Research Council within the European Union's Horizon 2020 program (grant agreement 951393).

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

Tortured Phrases as a Sign of Possible Misconduct in Proceedings From an Engineering Conference

Wendeline Swart,[1] Ophélie Fraisier-Vannier,[1] Guillaume Cabanac[1,2]

Objective Tortured phrases[1] in scientific articles, such as *amino corrosive* instead of *amino acid*, are potential signals of scientific misconduct. They alter established expressions using synonyms to avoid plagiarism detection. We highlighted the presence of tortured phrases in 1323 conference proceedings and identified the most impacted conference.

Design We used the Problematic Paper Screener (PPS),[2] a tool we developed to detect research integrity issues, to identify conference proceedings published containing tortured articles, defined as articles with 5 or more tortured phrases or fewer than 5 tortured phrases but assessed as

Table 25-0986. Duplicated Abstract Findings

Publication	ISSN	DOIs, No.	DOIs for which duplicated abstract exists, No.	Duplicates without preprint, No.	Duplicates with preprint, No.	Differences in authorship, No.	Duplicates all journals, No.
Science	0036-8075	1642	17	0	17	7	0
PLOS One	1932-6203	9951	375	4	371	89	3
International Journal of Molecular Sciences	1422-0067	9688	374	2	372	23	1
Scientific Reports	2045-2322	9943	781	6	775	85	5
BioMed Research International	2314-6141	167	4	2	2	2	2
IGI Global	NA	19,088	196	188	8	47	0

Abbreviations: DOI, digital object identifier; NA, not applicable.

problematic by a human. We selected the conference with the largest number of tortured articles and probed with Dimensions (a bibliometric platform) and PubPeer (a postpublication peer review platform) the production of the experts listed in its proceedings. We counted the experts' numbers of retractions, expressions of concern, and other red flags.

Results The PPS tabulated 20,066 tortured articles published from 2005 to 2024 by 344 publishers. The conference proceedings subset comprised 7518 articles (37%) mainly published by the Institute of Electrical and Electronics Engineers (IEEE) (6156 [82%]; presented in 1348 conferences), AIP Publishing (648 [9%]), IOP Publishing (364 [5%]), and 12 other publishers.[3] The most problematic conference was the IEEE International Conference on Optimization, Computing, and Wireless Communication held in Ethiopia in January 2024, featuring 324 peer-reviewed articles affiliated with India and China only. We identified problematic content in the proceedings using 2 clues: (1) nonsensical text, found in 260 articles (80%) with a mean 11 tortured phrases (**Figure 25-1029**), and (2) questionable expertise in 2 listed groups. The organizing committee (34 individuals) included 6 persons whose work had been commented on on PubPeer for various concerns on 31 articles, 21 of which have been retracted. Concerns included tortured phrases, suspicion of purchase of an authorship position, plagiarized sections, nonsensical equations, and irrelevant references. The reviewers (134 individuals) included 31 persons (23%) who authored 67 articles that were commented on on PubPeer for integrity issues, 8 of which had an expression of concern and 9 of which have been retracted. Eight reviewers have had several retractions, including 1 reviewer with 3 retractions due to violations of IEEE policy on authorship.

Conclusions As of February 2025, IEEE tagged with an expression of concern 15 of 260 articles (6%) flagged by the PPS (with a mean of 4 tortured phrases per article) and reported on PubPeer. IEEE had not issued expressions of concern or other corrective labels for articles with more tortured phrases. Limitations of this analysis include the focus on a single conference and the possibility that the PPS may identify false positives or less serious problems that do not require an expression of concern or retraction. However, reassessment may be needed for many of the 20,066 PPS-flagged articles.

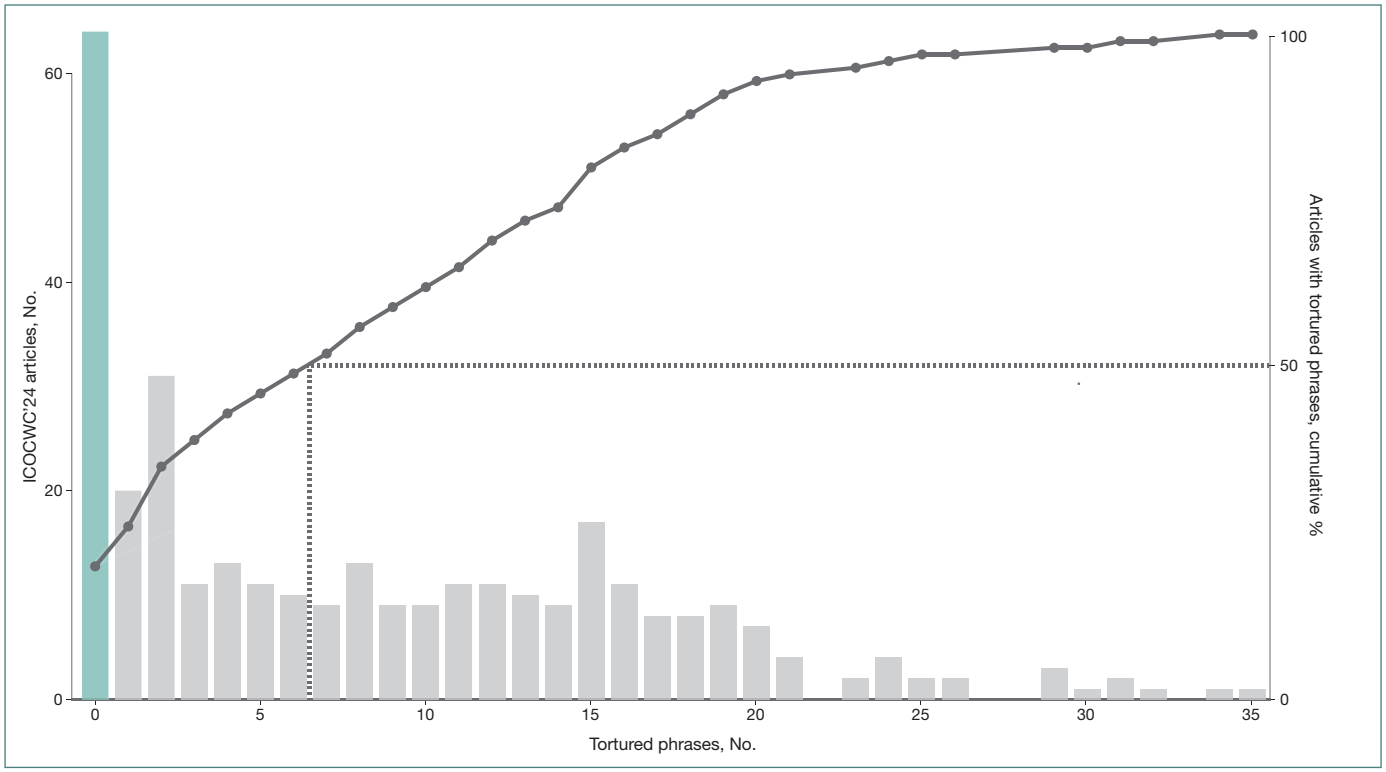
References

1. Cabanac G, Labbé C, Magazinov A. Tortured phrases: a dubious writing style emerging in science—evidence of critical issues affecting established journals. *arXiv*. Preprint posted online July 12, 2021. doi:10.48550/arXiv.2107.06751

2. Cabanac G, Labbé C, Magazinov A. The 'Problematic Paper Screener' automatically selects suspect publications for post-publication (re)assessment. *arXiv*. Preprint posted online October 7, 2022. doi:10.48550/arXiv.2210.04895

3. Swart W, Cabanac G. Year after year: Tortured conference series thriving in computer science. *arXiv*. Preprint posted online October 17, 2023. doi:10.48550/arXiv.2401.02422

Figure 25-1029. Pareto Chart of the Number of Tortured Phrases per Article Published in the Proceedings of the 2024 Institute of Electrical and Electronic Engineers International Conference on Optimization, Computing, and Wireless Communication (ICOCWC'24)



50% of the 324 ICOCWC'24 articles contained ≥7 tortured phrases. Bars represent the number of articles; trend line represents the cumulative percentage of articles with tortured phrases.

[1]Université de Toulouse, IRIT (UMR 5505 CNRS), Toulouse, France, guillaume.cabanac@univ-tlse3.fr; [2]Institut Universitaire de France, Paris, France.

Conflict of Interest Disclosures Guillaume Cabanac is the administrator of the Problematic Paper Screener, a public platform that uses metadata from Digital Science and PubPeer via no-cost agreements, and has been in touch with most of the major publishers and their integrity officers, offering pro bono consulting regarding detection tools to various actors in the field, including Clear Skies, Morressier, River Valley, Signals, and STM. No other disclosures were reported.

Funding/Support Guillaume Cabanac received funding from the Institut Universitaire de France and the NanoBubbles project, which received Synergy grant funding from the European Research Council within the European Union's Horizon 2020 program (grant agreement No. 951393).

Virtual

Plagiarism and Publication Fraud Revealed by Dissernet

Larisa Melikhova,[1] Andrei Rostovtsev,[2] Vasiliy Vlassov[3]

Objective Dissernet is a volunteer network combating plagiarism since 2013. Since 2020, we expanded our assessment to translation plagiarism (ie, using another author's text translated into a different language) from and into Russian and to purchasing of authorship. This report covers progress since 2016.

Design In 2020, by tracing authors known for earlier research misconduct, we located potential international predatory journals and checked their publications for plagiarism. All known predatory journals and potential predatory journals were screened for compliance with 2 indicators: explosive growth of publications by Russian authors and the fake collaboration index. We selected predatory journals that were indexed in Scopus and Web of Science (WoS) and examined whether they contained plagiarism translated from Russian-language sources. These data were assessed and presented in a report by the Commission for Counteracting the Falsification of Scientific Research, which was appointed by the Russian Academy of Sciences.[1] These journals have been expelled from Scopus and WoS. This work was continued by studying translation plagiarism from and into Ukrainian language (in collaboration with Ukrainian scientists). In 2021 to 2022, Dissernet detected another mass fraud: purchased coauthorship.[2] We downloaded a database of the offers at the International Publisher website and identified publications in journals indexed in WoS and/or Scopus with abstracts that coincided with offers in the database.

Results In the initial search, we selected 94 predatory journals, 83 of which were indexed in Scopus and 18 in WoS, and found in them 259 publications containing plagiarism translated from Russian-language sources. In examining purchased coauthorship, we identified 418 publications in journals indexed in Scopus and/or WoS that coincided with offers from International Publisher, at least 33 of which were later retracted by the journals. By May 2025, 5962 journal articles containing evidence of various types of research misconduct had been identified on the Dissernet website.[3] A total of 837 authors had published articles with translation plagiarism; the most common disciplines included economics, law, and education. We detected 842 authors who bought authorship through International Publisher[3]; the most common disciplines included medical sciences, technical sciences, economics, and law. Authors of publications with different types of violations identified by Dissernet (N = 2880) were predominantly from Russia: 2418 people (84%). Among the remaining 462 authors, 184 (40%) were from Eastern countries (Iran, Iraq, India, China, etc) and 157 (34%) were from countries of post-Soviet space (Kazakhstan, Tajikistan, Ukraine, etc). All results of the work described herein are presented on the Dissernet site (in Russian).[3] The publications with academic misconduct can be seen in the sections "Journals" (referring to a particular journal) and "Persons" (referring to authors of the articles).

Conclusions The practice of translation plagiarism is much easier today due to the use of artificial intelligence. Purchasing coauthorship became popular at the international level. We consider these findings as a further step in exploration of translation plagiarism in the countries of the post-Soviet space and beyond.

References

1. Chawla DS. Top officials at Russian universities embroiled in plagiarism scandal. August 14, 2020. Nature Index. Accessed July 16, 2025. https://www.nature.com/nature-index/news/top-officials-rectors-russian-universities-embroiled-plagiarism-scandal

2. Rostovtsev A, Melikhova L. A disaster that has become normal. *TRV Science*. 2023;370:1, 3. Accessed July 16, 2025. https://www.trv-science.ru/2023/01/katastrofa-kotoraya-stala-normoj/

3. Dissernet. Accessed July 16, 2025. https://dissernet.org/

[1]Dissernet, Netanya, Israel, larisamelikhova@gmail.com; [2]Dissernet, Budva, Montenegro; [3]International Health Equity Agency, Tel Aviv, Israel.

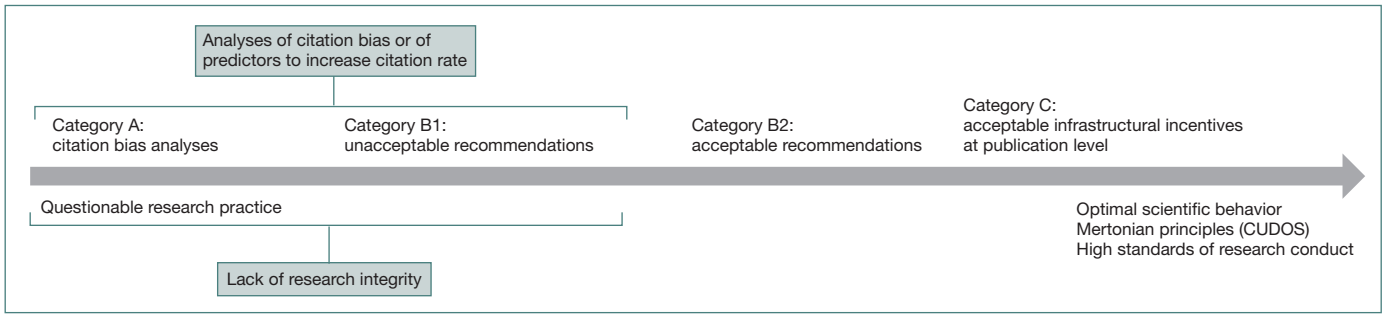
Conflict of Interest Disclosures None reported.

Citation Biases and Citation-Boosting Strategies: A Scoping Review of Predictors

Hans Lund,[1] Karen Lie,[2] Karen Robinson,[3] Jong-Wook Ban,[4] Birgitte Nørgaard[5]

Objective Researchers' citation practices are influenced by a wide range of factors, many of which are unrelated to scientific quality or relevance, raising concerns about citation bias. Given the conceptual ambiguity, methodological heterogeneity, and ethical implications surrounding this topic, a scoping review was conducted to map the literature, classify study intentions, and identify research gaps. The objective was to systematically map the predictors of citation

Figure 25-1083. Continuum of Citation Rate Predictors



We made no attempt to make a final list of categories B1 and B2 separately owing to a lack of clear criteria for defining each category. CUDOS indicates communalism, universalism, disinterestedness, and organized skepticism.

rates and categorize the types of studies evaluating these factors.

Design This was a scoping review to identify all studies assessing predictors of citation rates within the health sciences, with a particular focus on citation bias—defined as the selective citation of literature based on specific characteristics rather than scientific merit.

Results A total of 165 studies published between 1982 and 2023 were included. Fifty-four distinct factors were identified, encompassing 4,471,352 original studies, and were grouped into 4 categories: author-related (n = 17), study-related (n = 19), reporting-related (n = 10), and journal-related (n = 8). Among the 786 analyses evaluating these factors, 429 (54.6%) identified a significant increase in citation rates, 49 (6.2%) found a decrease, and 251 (31.9%) reported no effect. Most factors associated with increased citation rates were unrelated to the scientific content of the studies and instead reflected extrinsic characteristics. As the intention, interpretation of results, and conclusions varied considerably across the included studies, we categorized them based on how they analyzed factors associated with citation rates (**Figure 25-1083**). Of the 165 studies, 77 (46.7%) were classified under category A, which examined citation bias. In contrast, 81 studies (49.1%) (categories B1 and B2) recommended leveraging specific characteristics to enhance authors' own citation rates. Category B was further subdivided due to ambiguity in distinguishing ethically acceptable recommendations from potentially conflicting (ie, unacceptable) ones. Seven studies (4.2%) examined factors that could improve the accessibility of published studies. **Figure 25-1083** presents these categories along a continuum ranging from questionable research practices to optimal scientific behavior.

Conclusions Many factors associated with increased citation rates were unrelated to the scientific content of the studies. Furthermore, authors of nearly half of the studies explicitly recommended modifying paper characteristics to boost citations rather than prioritizing scientific contribution. Such recommendations may conflict with principles of scientific integrity, which emphasize relevance and methodological rigor over strategic citation practices. Furthermore, the results provide a robust foundation for conducting meta-analyses to quantify the extent of the problem, as they offer a comprehensive and systematically derived list of factors influencing citation rates. The high proportion of significant results suggests possible publication bias.

[1]Section of Evidence-Based Practice, Western Norway University of Applied Sciences, Bergen, Norway, hans.lund@hvl.no; [2]VID Specialized University, Bergen, Norway; [3]Johns Hopkins Evidence-Based Practice Center, Johns Hopkins University, Baltimore, MD, US; [4]Section of Hospital Medicine, University of Chicago, Chicago, IL, US; [5]Department of Public Health, University of Southern Denmark, Odense, Denmark.

Conflict of Interest Disclosures: None reported.

Open and Public Access

Virtual

Assessing the Cost-Effectiveness of Open Access Publishing in Dermatology Journals

Dante Dahabreh,[1] Kenny T. L. Ta,[2] Angela Rose Loczi-Storm,[3] Olivia V. Lim,[4] Dana Chen,[5] Tasneem M. Y. Issa,[6] Alexandria I. Kristensen-Cabrera,[6] Rahib K. Islam,[7] Robert P. Dellavalle,[6,8] Eamonn Maher[6,8]

Objective Open access (OA) publishing has revolutionized academic dissemination by providing free research access. However, the financial implications for authors and the impact on journal prestige, particularly in dermatology, are often unclear. This lack of clarity complicates researchers' publication decisions. This study aimed to evaluate the cost-effectiveness of OA publication models in dermatology journals.

Design We conducted a cross-sectional study of 144 dermatology journals from the Scimago database between May 13 and June 16, 2024. Of these, 106 journals met inclusion criteria (indexed in PubMed and having an OA option); 38 were excluded for not meeting inclusion criteria, having nonfunctional websites, or missing citation metrics. Journals were categorized as hybrid or full OA. We extracted 2022 metrics (CiteScore, h-index, Scimago journal rank, and 2-year Impact Factor), averaging them for a composite impact score. Article processing charges (APCs) were collected from journal websites. Cost-effectiveness was calculated as impact score divided by total OA publishing cost (cost for free OA journals was set to 1). Statistical analyses, including Wilcoxon

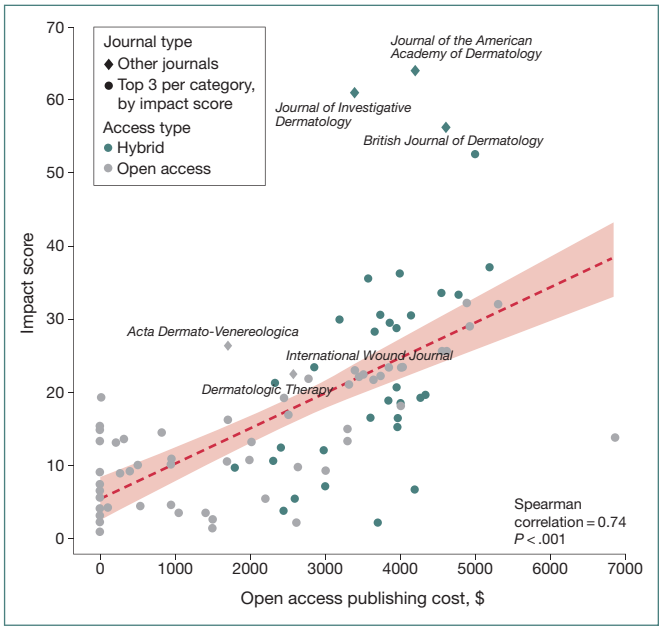
rank sum test and Spearman correlation, were performed using R software (R Foundation for Statistical Computing).

Results Median OA publishing costs were significantly higher for hybrid journals than full OA journals ($3675 vs $350, respectively; $P < .05$). Hybrid journals also showed significantly higher median impact scores than full OA journals (23.95 vs 9.23; $P < .05$) (**Figure 24-0826**). A moderate positive correlation ($R = 0.66$) was observed between OA publishing costs and impact scores. The *Journal of Drugs in Dermatology*, *Indian Journal of Dermatology*, and *Dermatology Online Journal* were the top 3 cost-effective paid OA options. The *Journal of Clinical and Aesthetic Dermatology* and the *Indian Journal of Dermatology, Venereology, and Leprology* were identified as top free OA journals.

Conclusions Our findings indicate a trade-off in dermatology publishing: hybrid journals, despite higher APCs, generally achieved greater impact scores, suggesting a balance between cost and prestige. While our impact score used established metrics, its broader validation and comparability across other medical subspecialties require further investigation. This study's cost-effectiveness calculations are based on listed APCs and do not account for fee-assistance programs (offered by 25 of 106 journals) or membership discounts (available for 22 of 106 journals), which can significantly alter the actual cost for individual authors. Future research, incorporating sensitivity analyses on varying APCs (including potential waivers or discounts) and exploring broader measures of influence, is essential to enhance the generalizability and impact of these findings and to better inform authors' complex OA publication decisions.

[1]Department of Internal Medicine, Kaiser Permanente, Oakland, CA, US; [2]University of Minnesota Medical School, Minneapolis, MN, US, ta000030@umn.edu; [3]College of Osteopathic Medicine of the Pacific Northwest, Western University of Health Sciences, Lebanon, OR, US; [4]Joan C. Edwards School of Medicine, Marshall University, Huntington, WV, US; [5]School of Osteopathic Medicine in Arizona, A.T. Still University, Mesa, AZ, US; [6]University of Minnesota Medical School, Minneapolis, MN, US; [7]Louisiana State University Health Sciences Center, New Orleans, LA, US; [8]Department of Dermatology, University of Minnesota, Minneapolis, MN, US.

Figure 24-0826. Impact Score vs Open Access Publishing Total Cost



Dashed line indicates the line of best fit, shaded area standard error

Conflict of Interest Disclosures None reported.

Additional Information ChatGPT Omega 3 was used on June 18, 2024, to help with statistical analysis and figure generation.

Public Access to Information Cited in Rare Disease Reports

Mengyuan Fu,[1,2] Kexin Ling,[3] Xinyi Zhou,[3] Sneha Dave,[4] Can Li,[3] Luwen Shi,[2,3] Xiaodong Guan,[2,3] Joseph S. Ross[1,5,6]

Objective This study aims to examine whether the information sources cited within National Organization for Rare Disorders (NORD) rare disease reports were publicly available, as well as any associated costs to purchase access, given that patients, caregivers, and clinicians may rely on these sources for critical information to inform their clinical decision-making.

Design We randomly selected 20% of all NORD rare disease reports available as of February 2024.[1] Citations within these reports were classified into 5 source types, including publications in scholarly journals (research articles, case reports, reviews), textbooks, and other internet resources. Each citation was searched using Google and categorized as open access (freely available online), accessible at cost, or unavailable online. For citations with associated costs to purchase access, we recorded the lowest listed price before taxes. A 10% cross-sample validation was conducted between each pair of researchers to check consistency. All data were collected between April and May, 2024. We used descriptive statistics to calculate the proportion of citations categorized by availability, overall and by year and source type, as well as median costs, overall and by source type. We used χ^2 tests to evaluate differences in open access rates by year and source type. A 2-sided $P < .05$ was considered statistically significant. Analyses were performed using Stata MP version 16.0.

Results Among 1333 NORD rare disease reports available as of February 2024, 267 were randomly selected for study. These 267 reports included 4445 citations to 3506 unique information sources; the median (IQR) number of citations per report was 15 (11-20), including 9 (5-13) research articles, 0 (0-1) case reports, 0 (0-0) review articles, 1 (0-3) textbook, and 2 (1-4) other internet resources. Among 3506 unique information sources cited, 1750 (49.9%) were open access, 1381 (39.4%) were accessible at cost, and 375 (10.7%) were unavailable online (**Table 25-0964**). Among citations accessible only at cost, the median (IQR) cost was $31.5 ($15.0-$40.0), whereas the total cost per report for all citations accessible only at cost was $167.7 ($82.0-$258.7). The overall proportion of citations published in scholarly journals available as open access increased from 30.0%

Table 25-0964. Public Availability of Citations Within 267 National Organization for Rare Disorders Reports

Characteristic	No. (%) [95% CI]					*P* value[a]
	Publications in scholarly journals			Textbook	Internet resource	
	Research article	Case report	Review			
Public availability						
Open access	1426 (52.0) [50.1-53.8]	113 (41.4) [35.5-47.2]	115 (46.2) [40.0-52.4]	83 (18.7) [15.1-22.3]	626 (85.2) [82.5-87.6]	<.001
Accessible at cost[b]	1090 (39.7) [37.9-41.6]	93 (34.1) [28.4-39.7]	113 (45.4) [39.2-51.6]	322 (72.5) [68.4-76.7]	28 (3.8) [2.4-5.2]	<.001
Pay-per-view access	22 (0.8) [0.5-1.1]	4 (1.5) [0.0-2.9]	1 (0.4) [0-1.2]	NA	NA	NA
Limited period of online access	645 (23.5) [21.9-25.1]	53 (19.4) [14.7-24.1]	49 (19.7) [14.7-24.6]	NA	NA	NA
Unlimited online-only access	16 (0.6) [0.3-0.9]	3 (1.1) [0-2.3]	0	NA	NA	NA
PDF download	325 (11.8) [10.6-13.1]	27 (9.9) [6.3-13.4]	54 (21.7) [16.6-26.8]	NA	NA	NA
Specific type not mentioned	82 (3.0) [2.4-3.6]	6 (2.2) [0.5-3.9]	9 (3.6) [1.3-5.9]	NA	NA	NA
Unavailable online	227 (8.3) [7.2-9.3]	67 (24.5) [19.4-29.6]	21 (8.4) [5.0-11.9]	39 (8.8) [6.2-11.4]	82 (11.2) [8.9-13.4]	<.001
Total No.	2743	273	249	444	736	NA
Cost of information sources accessible only at cost						
Cost, median (IQR), $[c]	31.5 (15.0-40.0)	31.5 (12.0-40.7)	36.0 (28.0-40.7)	30.8 (9-76.2)	23.0 (NC)	NA

Abbreviations: NA, not applicable; NC, not calculable.

[a]Calculated using χ^2 tests to compare the distribution of each citation availability category across source types.

[b]Each scholarly publication was assigned into 1 independent subcategory according to which access had the lowest cost.

[c]Costs were converted into US dollars uniformly on the basis of the exchange rate on May 24, 2024.

(258/860) prior to 2001 to 79.8% (95/119) from 2021 onwards ($P < .001$).

Conclusions While NORD should be applauded for making its rare disease reports publicly available, we found that half of the information sources cited by these reports were not open access, potentially limiting equitable access to critical information for rare disease communities.

Reference

1. National Organization for Rare Disorders. Rare disease database. Accessed March 1, 2024. https://rarediseases.org/rare-diseases/

[1]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, US, joseph.ross@yale.edu; [2]International Research Center for Medicinal Administration, Peking University, Beijing, China; [3]Department of Pharmacy Administration and Clinical Pharmacy, School of Pharmaceutical Sciences, Peking University, Beijing, China; [4]Generation Patient, Indianapolis, IN, US; [5]Center for Outcomes Research and Evaluation, Yale–New Haven Health System, New Haven, CT, US; [6]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, US.

Conflict of Interest Disclosures Mengyuan Fu currently receives research support from the National Natural Science Foundation of China (72304011). Sneha Dave is the founder and executive director of Generation Patient, an organization that receives grant funding from Arnold Ventures, Commonwealth Fund, Disability Inclusion Fund, Third Wave Fund, Responsible Technology Youth Power Fund, the Helmsley Charitable Trust, Patient-Centered Outcomes Research Institute, and Lucile Packard Foundation for Children's Health. Joseph S. Ross currently receives research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, 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), and from Arnold Ventures; formerly received research support from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology (NEST) and from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644); was 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. that was settled in September 2022; and is a deputy editor for *JAMA*. No other disclosures were reported.

Open Science

In-person

Citations of Articles With Open Science Indicators in the French Open Science Monitor Dataset

Giovanni Colavizza,[1,2] Lauren Cadwallader,[3] Iain Hrynaszkiewicz[3]

Objective This research explored whether articles that demonstrate 1 or more open science practices received more citations on average than similar articles that do not demonstrate open science practices by applying an existing linear regression model[1] to an openly available dataset.

Design This study applied a previously published citation prediction model[1] with data from openly available sources including OpenAlex, DataCite, CrossRef, and the French Open Science Monitor dataset,[2] which includes all research publications from open data sources with at least 1 French author. This study evaluated whether open science practices in publications are associated with changes in citation impact. A total of 479,700 journal articles were analyzed, which were from across 10 broad categories of research domains, have at least 1 French author, were published in 2020-2022, and for which the metadata, including citation counts, could be

extracted from the publicly available sources at the time of this study (July 2024). The citation prediction method controlled for factors known to influence citations such as article age, open access status, research discipline, length of reference lists, and numbers of authors. Associations were assessed with an ordinary least squares model.

Results The analysis showed that a citation advantage was associated with all 3 open science practices (data, code, and preprint sharing), both individually and cumulatively. Posting a preprint had the largest association with citations and was correlated with a 19.0% higher number of citations, compared with articles that did not have preprints. The associations with sharing research data and code were smaller—and correlated with 14.3% and 13.5% higher number of citations, respectively. These advantages were cumulative, that is, articles with a preprint, shared research data, and shared code attracted a mean of 46.8% more citations than articles that do not share these outputs. Variation in citation advantage was seen across disciplines (**Table 25-0867**). For example, articles that had preprint postings had a 1.0% higher number of citations in the fields of earth, ecology, energy and 62.5% more citations in the humanities, compared with articles in the same disciplines that did not have preprints. Study limitations include the dataset bias toward French publications (we might observe different citation correlations in a more international sample), and the results for the data and code sharing citation advantage were not directly comparable with the results in previous studies due to differences in the definition and calculation of open science practices.

Conclusion This study explored the French Open Science Monitor dataset and investigated whether open science indicators were associated with a citation premium received by the publications that follow them. To our knowledge, this is the largest scale analysis of such a question to date using nearly 500,000 publications from many scientific domains, which found a positive association between the open science practices of data, code, and preprint sharing and postpublication citation counts.

References

1. Colavizza G, Cadwallader L, LaFlamme M, et al. An analysis of the effects of sharing research data, code, and preprints on citations. *PLOS One*. 2024;19(10): e0311493. doi:10.1371/journal.pone.0311493

2. Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. Baromètre de la science ouverte (général). Accessed July 15, 2024. https://data.enseignementsup-recherche.gouv.fr/explore/dataset/open-access-monitor-france/information

[1]University of Bologna, Bologna, Italy; [2]University of Copenhagen, Copenhagen, Denmark; [3]PLOS, Cambridge, UK, lcadwallader@plos.org.

Conflict of Interest Disclosures PLOS works to promote adoption of open science practices and provided funding for this study for the collection, management, analysis, and interpretation of the data by Giovanni Colavizza. PLOS also provided support in the form of salaries for authors Lauren Cadwallader and Iain Hrynaszkiewicz. PLOS also had a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the abstract; and decision to submit the abstract for presentation.

Additional Information Data and code necessary to reproduce the results are available: https://doi.org/10.6084/m9.figshare.27822663

Automated Interpretation of Statistical Tables to Assess Reporting Errors and Associations With Open Science Policies in Economics Journals

Stephan B. Bruns,[1] Helmut Herwartz,[2] John P. A. Ioannidis,[3] Chris-Gabriel Islam,[2] Fabian H. C. Raters[2]

Objective The economics literature predominantly reports statistical values in data tables. This study developed a tool for extraction of statistical values in tables to identify reporting errors and to analyze the association between

Table 25-0867. Journal Articles by Discipline, Open Science Practice, and Citation Advantage

Discipline	No. of articles with preprint match	No. of articles with data shared	No. of articles with software shared	Change in citations with preprint match, %	Change in citations with data sharing, %	Change in citations with software sharing, %
Computer and information sciences (n = 17,944)	1612	1989	812	22.0	6.5	5.3
Biology (n = 81,063)	8258	8341	2296	25.3	18.8	11.4
Chemistry (n = 26,698)	1405	1432	176	14.5	6.4	33.0
Medical research (n = 145,846)	4474	4870	768	17.7	34.9	5.7
Physical sciences, astronomy (n = 41,801)	15,099	2638	527	24.9	10.4	21.3
Humanities (n = 37,668)	526	1166	153	62.5	–1.4	15.0
Social sciences (n = 34,565)	534	1125	135	–3.8	2.3	38.0
Earth, ecology, energy and applied biology (n = 43,165)	6665	6949	1335	1.0	12.0	17.9
Mathematics (n = 15,109)	3245	489	264	9.7	12.9	12.1
Engineering (n = 26,573)	1232	976	263	23.1	3.2	16.6

journal data and code availability policies and article characteristics.

Design DORIS (Diagnosis of Reporting Errors in Scraped Tables)[1,2] automatically extracts statistical values (eg, coefficients and SEs) from tables via web scraping in R to extract tables from HTML articles and text mining in Python to interpret data. This analysis included the top 50 economics journals that provided articles in HTML from 1998 to 2016. Tests were divided into 2 categories: main (first 3 table rows from main tables) and nonmain (all other rows from main tables and other tables). We used a staggered difference-in-differences design to assess the association between data and code availability policies and article outcomes[3]: reporting errors using a dummy for inconsistency altering significance level; statistical significance proxied by *z*-value, presence of visual indicators of statistical significance, and type of statistical value reported; logarithmized number of tests per article, tables with tests, and a dummy for appendix position; and citations measured by the logarithmized citation count for the first 7 years.

Results The study included 578,132 statistical tests (median: 27 per table) from 15,725 tables (median: 4 per article) in 3746 articles (median: 76 per journal) that were published in 31 journals. Based on a sample of 3068 statistical tests, DORIS had a false discovery rate of 1.2%. Analysis revealed that 547 of 3677 articles (14.8%) had at least 1 reporting error in main tests altering the statistical significance of findings. The errors were prevalently oriented toward lowering *P* values and claiming statistical significance. Assessment of data and code availability policies included 21 journals and 535,838 tests. These policies were associated with slightly less emphasis on statistical significance (smaller z-values: −0.16 [90% CI, −0.33 to 0], *P* = .10; fewer statistical significance symbols: −4 [90% CI, −6 to −2] percentage points [pp]; *P* = .004; fewer tests reported with a focus on significance (*t*, *z*, or *P* values instead of SEs and CIs): −8 [90% CI, −13 to −3] pp; *P* = .004), more rigor in reporting (more tests: 11% [90% CI, −1% to 22%]; *P* = .12; more tables: 15% [90% CI, 7% to 23%]; *P* = .002; more appendix tables: 4 [90% CI, 2 to 7] pp; *P* = .009), and fewer reporting errors in nonmain tests (−0.2 [90% CI, −0.4 to 0] pp; *P* = .08), but there was no difference in citations (−9% [90% CI, −21% to 4%]; *P* = .25).

Conclusions DORIS may be used in the peer review process in the economics literature to improve article quality and to generate large scale data for future meta-research. Data and code availability policies may help improve the reliability of published economics research.

References

1. Better Papers. Home page. Accessed June 12, 2025. https://betterpapers.org/#sec-faqs

2. Bruns SB, Herwartz H, Ioannidis JPA, Islam C-G, Raters FHC. Statistical reporting errors in economics. *MetaArXiv*. Preprint posted online September 01, 2023. doi:10.31222/osf.io/mbx62

3. Borusyak, K, Jaravel, X, Spiess, J. Revisiting event-study designs: robust and efficient estimation. *Rev Econ Studies*. 2024;91(6):3253-3285. doi:10.1093/restud/rdae007

[1]Hasselt University, Hasselt, Belgium, stephan.b.bruns@gmail.com; [2]Georg August University Göttingen, Göttingen, Germany; [3]Meta-Research Innovation Center at Stanford (METRICS), Stanford, CA, US.

Conflict of Interest Disclosures John P. A. Ioannidis 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 Funding was received from the German Research Foundation (DFG) under the project "Replications in Empirical Economics: Necessity, Incentives and Impact," with follow up project "Selective Reporting and the Evolving Research Landscape in Economics" (project number: 405039391) and from Hasselt University within the framework of BOF BILA (BOF21BL08).

Role of Funder/Sponsor DFG and BOF BILA funding helped 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.

Detection of Open Science Practices in Major Medical Journals: A Survey and Diagnostic Accuracy of Automatic Tools Using Sensitivity and Specificity

Constant Vinatier,[1] Ayu Putu Madri Dewi,[2] Gwénaël Dumont,[1] Tracey Weissgerber,[3] Vladislav Nachev,[3] Gowri Gopalakrishna,[2,3,4] Maud Scheidecker,[1] François-Joseph Arnault,[1] Nicholas J. DeVito,[5] Guillaume Freyermuth,[6] Mathieu Acher,[6,10] Gauthier Le Bartz Lyan,[6] Inge Stegeman,[7,8] Mariska M. G. Leeflang,[2] F. Naudet[9,10]

Objective Despite open science policies in major biomedical journals, adherence remains uncertain. This study evaluated automated tools, from regular expressions to large language models (LLMs), for assessing core open science practices in leading biomedical journals.

Design We retrospectively assessed research articles from a sample of 10 major generalist medical journals (*Annals of-Internal Medicine*, *BMJ*, *BMC Medicine*, *Canadian Medical Association Journal* [*CMAJ*], *JAMA*, *JAMA Network Open*, *Lancet*, *Nature Medicine*, *New England Journal of Medicine*, and *PLoS Medicine*) from 2020 to 2023. Articles were retrieved via PubMed using a Peer Review of Electronic Search Strategies (PRESS) search strategy. The database comprised random samples of 103 randomized controlled trials (RCTs), 98 meta-analyses (MAs), and 111 other research articles (RAs). We evaluated 13 open science practices, including study registration, data sharing, and protocol sharing (open access or upon request). Each article was evaluated by 2 independent raters, with any disagreements resolved by a third rater. Seven different automated tools—rtransparent, oddpub, ctRegistries, ContriBot, DataSeer, SciScore, and an LLM (Llama 3-70B)—were used. Diagnostic

accuracies were estimated using sensitivities, specificities, F1 scores, and LR+ and LR-.

Results Manual extraction in the 312 articles identified registration in 98% (101/103) of RCTs, 69% (68/98) of MAs, and 18% (20/111) of RAs. Open data were present in 6% (6/103) of RCTs, 36% (35/98) of MAs, and 13% (15/111) of RAs and accessible upon request in 78% (80/103), 41% (40/98), and 59% (66/111), respectively. Protocols were openly available in 84% (87/103) of RCTs, 64% (63/98) of MAs, and 20% (22/111) of RAs and accessible upon request in 5% (5/103), 3% (3/98), and 3% (3/111), respectively. The accuracy of automated tools varied depending on the practice evaluated, with F1 scores ranging from 1.00 (Conflict of Interest statement, rtransparent) to 0.16 (SciScore, registration). For study registration, a simple tool using regular expressions, such as rtransparent, demonstrated good sensitivity (77%; 95% CI, 70%-83%) and high specificity (93%; 95% CI, 88%-97%). Data sharing detection remained challenging; for instance, rtransparent detects data sharing with a sensitivity of 74% (95% CI, 68%-80%) and a specificity of 59% (95% CI, 46%-70%). Different diagnostic accuracies were observed depending on the type of research and the journal, likely due to different formatting standards. All results are shown in **Table 25-1025**. Limitations include the declarative nature of some practices (eg, data sharing).

Conclusions Our study provides a detailed description of core open science practices across leading biomedical journals. It also highlights current challenges regarding the accuracy of automated tools in detecting these practices. While these tools likely provide valuable insights into overall practices, it is crucial to remain aware of the potential ranking biases introduced by these tools, as well as their limitations in providing detailed feedback for individual studies.

[1]Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail), UMR_S 1085, Rennes, France, constant.vinatier1@gmail.com; [2]Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, the Netherlands; [3]QUEST Center for Responsible Research, Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Berlin, Germany; [4]Department of Epidemiology, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, the Netherlands; [5]Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK; [6]Univ Rennes, IRISA, Inria, CNRS, Rennes, France; [7]Department of Otorhinolaryngology and Head and Neck Surgery, University Medical Center Utrecht, Utrecht, the Netherlands; [8]Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands; [9]Univ Rennes, CHU Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail), UMR_S 1085, Rennes, France; [10]Institut Universitaire de France (IUF), France.

Conflict of Interest Disclosures None reported.

Funding/Support As part of the OSIRIS project, this work was supported by the European Union's Horizon Europe Research and Innovation Program under grant agreement number 101094725. Constant Vinatier, Ayu Putu Madri Dewi, Gowri Gopalakrishna, Nicholas J. DeVito, Inge Stegeman, Mariska M. G. Leeflang, and F. Naudet are members of this project.

Table 25-1025. Description of the Accuracy of Automatic Tools

Tool	Measure	Registration (189/312 [61%])	Data sharing (242/312 [78%])	Code sharing (75/312 [24%])	COI statement (312/312 [100%])
oddpub	Sensitivity (95% CI)	NA	0.70 (0.64-0.76)	0.72 (0.56-0.85)	NA
	Specificity (95% CI)	NA	0.61 (0.49-0.73)	0.88 (0.84-0.92)	NA
	F1 score	NA	0.77	0.60	NA
	LR+	NA	1.81 (1.33-2.46)	6.36 (4.25-9.52)	NA
	LR-	NA	0.49 (0.38-0.64)	0.31 (0.19-0.51)	NA
rtransparent	Sensitivity (95% CI)	0.77 (0.70-0.83)	0.74 (0.68-0.80)	0.68 (0.51-0.81)	1.00 (0.99-1.00)
	Specificity (95% CI)	0.93 (0.88-0.97)	0.59 (0.46-0.70)	0.96 (0.92-0.98)	NaN (0-1.00)
	F1 score	0.85	0.80	0.70	1.00
	LR+	11.80 (6.01-23.16)	1.80 (1.35-2.40)	16.00 (8.41-30.45)	NaN (NaN-NaN)
	LR-	0.25 (0.19-0.32)	0.44 (0.33-0.59)	0.34 (0.22-0.53)	NaN (NaN-NaN)
ctregistries	Sensitivity (95% CI)	0.57 (0.49-0.64)	NA	NA	NA
	Specificity (95% CI)	0.84 (0.76-0.90)	NA	NA	NA
	F1 score	0.68	NA	NA	NA
	LR+	3.48 (2.29-5.30)	NA	NA	NA
	LR-	0.52 (0.43-0.62)	NA	NA	NA
TRN screener	Sensitivity (95% CI)	0.57 (0.49-0.64)	NA	NA	NA
	Specificity (95% CI)	0.84 (0.76-0.90)	NA	NA	NA
	F1 score	0.68	NA	NA	NA
	LR+	3.48 (2.29-5.30)	NA	NA	NA
	LR-	0.52 (0.43-0.62)	NA	NA	NA
Contribot	Sensitivity (95% CI)	NA	NA	NA	NA
	Specificity (95% CI)	NA	NA	NA	NA
	F1 score	NA	NA	NA	NA
	LR+	NA	NA	NA	NA
	LR-	NA	NA	NA	NA
SciScore[a]	Sensitivity (95% CI)	0.09 (0.05-0.15)	0.41 (0.27-0.57)	0.19 (0.08-0.35)	NA
	Specificity (95% CI)	0.99 (0.94-1.00)	0.78 (0.64-0.88)	0.99 (0.97-1.00)	NA
	F1 score	0.17	0.50	0.31	NA
	LR+	8.80 (1.18-65.80)	1.88 (1.00-3.51)	33.11 (4.20-261.08)	NA
	LR-	0.92 (0.87-0.97)	0.75 (0.57-1.00)	0.82 (0.70-0.95)	NA
DataSeer[a]	Sensitivity (95% CI)	0.80 (0.72-0.85)	0.45 (0.30-0.61)	0.33 (0.18-0.52)	NA
	Specificity (95% CI)	0.93 (0.87-0.97)	0.86 (0.75-0.93)	0.99 (0.96-1.00)	NA
	F1 score	0.86	0.55	0.47826	NA
	LR+	11.81 (5.75-24.26)	3.23 (1.63-6.42)	33.67 (7.81-145.11)	NA
	LR-	0.22 (0.16-0.30)	0.63 (0.48-0.85)	0.67 (0.53-0.86)	NA
Llama3-70B	Sensitivity (95% CI)	0.55 (0.48-0.62)	0.78 (0.72-0.83)	0.64 (0.52-0.75)	0.74 (0.69-0.79)
	Specificity (95% CI)	0.95 (0.90-0.98)	0.66 (0.53-0.77)	0.74 (0.68-0.79)	NaN (0-1.00)
	F1 score	0.696	0.828	0.519	0.851
	LR+	11.28 (5.11-24.88)	2.27 (1.63-3.16)	2.45 (1.86-3.21)	NaN (NaN-NaN)
	LR-	0.47 (0.40-0.56)	0.34 (0.25-0.45)	0.49 (0.36-0.67)	NaN (NaN-NaN)

Abbreviations: COI, conflict of interest; NA, not applicable; NaN, not a number.
[a]This tool is part of a commercial company.

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

Pandemic Science

In-person

Analysis of Editorials on the Response to the H1N1 and COVID-19 Pandemics

Luka Ursić,[1,2] Nensi Bralić,[1] Giovanni Spitale,[2] Federico Germani,[2] Ana Marušić[1]

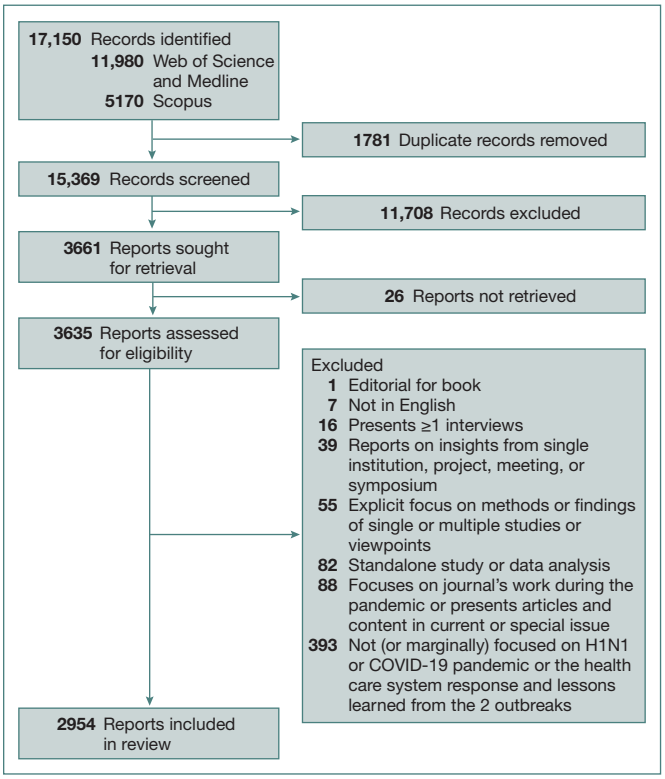
Objective Divisive discourse observed in news coverage or social media posts by researchers and the general public during the H1N1 and COVID-19 pandemics contributed to polarization and politicization, rapid shifts in sentiment, and concerning rates of misinformation.[1-3] We aimed to study the sentiment and natural language processing (NLP) analysis of editorial material in medical journals discussing the health care system response to the 2 pandemics.

Design Using a systematic review design, we searched Web of Science and PubMed up to May 22, 2024, for editorials, viewpoints, and similar opinion pieces discussing the following segments of the health care system response to the 2 pandemics: nonpharmaceutical interventions; misinformation or disinformation; health care resource allocation and management; health care system preparedness; mandates, policies, and guidance related to research, education, or new technologies; and lessons learned. We limited our search to 2009 to 2014 for the H1N1 pandemic and 2019 to 2024 for the COVID-19 pandemic. Three researchers screened the titles, abstracts, and full texts of the retrieved records for eligibility in pairs, resolving discrepancies through discussion. We extracted the text from the included records' PDFs using Python and performed a sentiment analysis using the Language Inquiry and Word Count 2022 software. We also used NLP approaches (flat lemmatization, convoluted lemmatization, rule-based autocoding, and cosine similarity) to determine the most frequently used lemmas in the 2 groups of editorials and explore topics based on their co-occurrence. We conducted inferential analyses at the level of the full sample and across specific categories (ie, stratified by pandemic, author country, and context of response). While we could not set a hypothesis for the exploratory NLP analysis, we hypothesized that in terms of sentiment, the editorials published during the COVID-19 pandemic would have higher scores for negative tone and emotion, personal and person-centered language, certitude, and all-or-none thinking and lower scores for analytical thinking.

Results Following the screening process (**Figure 25-0979**), we included 2954 editorials in the final dataset. We performed a preliminary analysis of 200 editorials: 25 for the H1N1 pandemic and 175 for the COVID-19 pandemic. In contrast to our initial hypothesis, the editorials for the H1N1 pandemic had higher values for negative tone (median, 2.21 [IQR, 1.74-3.07] vs 1.69 [IQR, 1.22-2.31]; $P = .007$) and certitude (median, 0.39 [IQR, 0.16-0.66] vs 0.23 [IQR, 0.12-0.37]; $P = .01$). There were no differences in negative emotion, analytical thinking, or all-or-none thinking.

Conclusions In this preliminary analysis, the editorials for the H1N1 pandemic contained more words connoting negative tone and certitude than those for the COVID-19 pandemic. These findings suggest that the researchers authoring these editorials expressed a more negative sentiment toward the public health response to the H1N1 pandemic and that they were seemingly more certain about their position.

Figure 25-0979. PRISMA Diagram



[a]Includes cases where the editorial overtly focused on impact or or response from the context of a single discipline or specialty or the response to a single event.

References

1. Schmidt H. Pandemics and politics: analyzing the politicization and polarization of pandemic-related reporting. *Newsp Res J*. 2023;44(1):26-52. doi:10.1177/07395329221095850

2. Lwin MO, Lu J, Sheldenkar A, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. *JMIR Public Health Surveill*. 2020;6(2):e19447. doi:10.2196/19447

3. Sule S, DaCosta MC, DeCou E, Gilson C, Wallace K, Goff SL. Communication of COVID-19 misinformation on social media by physicians in the US. *JAMA Netw Open*. 2023;6(8):e2328928. doi:10.1001/jamanetworkopen.2023.28928

[1]Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia, luka.ursic@mefst.hr;

[2]Institute of Biomedical Ethics and History of Medicine, University of Zürich, Zürich, Switzerland.

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 under grant agreement IP-2019-04-4882 and the University of Zürich Institute of Biomedical Ethics and History of Medicine under a Stehr-Boldt fellowship given to Luka Ursić.

Role of the Funder/Sponsor The funders had no role in the design of this study, its execution, analyses, interpretation of the data, or decision to submit results.

Consistency and Completeness of Retractions in Public Health Research on COVID-19

Caitlin J. Bakker,[1,2] Erin E. Reardon,[3] Sarah Jane Brown,[4] Nicole Theis-Mahon,[4] Sara Schroter,[5,6] Lex Bouter,[7,8] Maurice P. Zeegers[2]

Objective The scientific community responded rapidly to COVID-19, producing over 200,000 publications in 1 year.[1] This speed brought challenges, including a higher retraction rate.[2] While retraction helps correct the scientific record, retracted status may be inconsistently presented and notices may be incomplete.[3] During a pandemic, inconsistency and incompleteness can have immediate and long-term health impacts. We evaluated retraction presentation consistency and notice completeness for COVID-19 vs non–COVID-19 publications. We describe reasons for retraction and time from publication to retraction.

Design In March 2023, we retrieved retracted publications categorized as research articles or clinical studies in the subject area of public health and safety from Retraction Watch. A previous study focused on all retracted publications in this period[3]; this is a sub-study comparing COVID-19 with non–COVID-19 publications. Between April 28 and June 6, 2023, we assessed consistency in 11 databases (Academia.edu, CINAHL, Embase.com, Ovid Embase, Ovid Medline, PubMed, ResearchGate, SciHub, Scopus, Web of Science, and publisher websites) using 12 criteria from the International Committee of Medical Journal Editors and National Library of Medicine and notice completeness using 17 criteria from Retraction Watch and the Committee on Publication Ethics. Criteria were scored 0 if missing or 1 if present; partial scores were assigned when evaluating multicomponent criteria, such as bidirectional links. To ensure consistent scoring, a random subset of 21% (92 of 441) of retracted publications were independently reviewed by 2 researchers. Each researcher extracted data using Qualtrics forms, and scoring discrepancies were resolved by consensus. Following this calibration phase, remaining publications were extracted by a single reviewer. Kruskal-Wallis tests assessed differences in scores between COVID-19 and non–COVID-19 publications.

Results Of 441 publications, 47 were about COVID-19 and 394 were not. COVID-19 publications were published between 2019 and 2022, while non-COVID 19 publications were published between 1978 and 2022. COVID-19 publications were most frequently retracted due to concerns about reliability of data or results (15 [31.9%]) compared with plagiarism (82 [20.8%]) for non–COVID-19 publications. The median time between publication and retraction was 120 (IQR, 15-196) days for COVID-19 publications and 326 (IQR, 124-789) days for non–COVID-19 publications ($P < .001$). Across 11 databases, 41.2% (110 of 267) of records retrieved for retracted COVID-19 publications were marked as retracted compared with 47.6% (1225 of 2574) for non–COVID-19 publications. There was no statistically significant difference between consistency or completeness scores for COVID-19 vs non–COVID-19 retracted publications (**Table 25-1062**). No publications met all criteria.

Conclusions Incomplete and inconsistent information poses challenges for researchers and practitioners, undermining trust in scientific literature. We found no association between publications being about COVID-19 and the consistency or completeness of retraction information; however, publications about COVID-19 appeared to be retracted more quickly.

References

1. Shimray SR. Research done wrong: a comprehensive investigation of retracted publications in COVID-19. *Account Res*. 2022;30(7):393-406. doi:10.1080/08989621.2021.2014327

2. Yeo-Teh NSL, Tang BL. An alarming retraction rate for scientific publications on coronavirus disease 2019 (COVID-19). *Account Res*. 2020;28(1):47-53. doi:10.1080/08989621.2020.1782203

3. Bakker CJ, Reardon EE, Brown SJ, et al. Identification of retracted publications and completeness of retraction notices in public health. *J Clin Epidemiol*. 2024;173:111427. doi:10.1016/j.jclinepi.2024.111427

[1]University of Regina, Regina, Saskatchewan, Canada, caitlin.bakker@uregina.ca; [2]Maastricht University, Maastricht, the Netherlands; [3]Emory University, Atlanta, GA, US; [4]University of Minnesota, Minneapolis, MN, US; [5]*BMJ*, London, UK; [6]London School of Hygiene and Tropical Medicine, London, UK; [7]Amsterdam University Medical Center, Amsterdam, the Netherlands; [8]Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

Conflict of Interest Disclosures Caitlin J. Bakker is cochair of the National Information Standards Organization Communication of Retractions, Removals and Expressions of Concern Standing Committee. Lex Bouter 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 25-1062. Consistency Scores for Retracted Publications and Completeness Scores for Retraction Notices

Score type	Score, median (IQR)		*P* value
	COVID-19 publications	Non–COVID-19 publications	
Consistency	0.5 (0.5-4.0)	1.0 (1-4)	.53
Completeness	8.5 (6.0-10.5)	9.0 (7-10)	.13

Funding/Support This research is part of an ongoing PhD collaboration between *The BMJ (British Medical Journal)* and the team Meta-Research at Maastricht University (UM) on the responsible conduct of publishing scientific research. *The BMJ* is published by BMJ Group, a wholly owned subsidiary of the British Medical Association. UM is a public legal entity in the Netherlands. This study is part of Caitlin Bakker's self-funded BMJ/UM PhD. No exchange of funds has taken place for this research project.

Role of the Funder/Sponsor The authors are wholly responsible for 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 All authors express their own opinions and not necessarily that of their employers.

Paper Mills

In-person

Analysis of Cancer Research Discussion Text and References in High Impact Factor Journals for Possible Indicators of Paper Mills

Annie Whamond,[1] Adrian G. Barnett,[2] Jennifer A. Byrne[1,3]

Objective Paper mills are unethical organizations that provide low-value or fraudulent content to client authors.[1] Although image manipulation and incorrect reagent detection tools are available,[1] these approaches can require reviewers and readers to learn new skills. Scaled manuscript production through templates may also result in paper mill articles showing other common features. We aimed to help readers identify possible indicators of paper mill support by simply scanning article text and references. We focused on discussion sections for this study.

Design We used an exploratory, cross-sectional design[2] to develop a descriptive analysis of discussion text and references in high Impact Factor (IF) cancer journals, defined as IF of 7 or higher for journal categories oncology, biochemistry and molecular biology, or cell biology. We downloaded the Retraction Watch database on October 16, 2024, and filtered for retraction reason *paper mill*, subject *cancer*, and journals on our high IF list. To identify comparison article cohorts, we randomly sampled original cancer articles in (1) the same 8 journals as retracted articles or (2) 8 independent journals with no paper mill retractions and high IF maintained for 20 years or longer. For each article, we recorded the percentages of references first cited in the introduction, methods, results, or discussion section. Focusing on discussion sections, we classified each sentence as providing background, summary, comparison, interpretation or implication, limitation, or future direction[3] and recorded whether individual discussion sentences cited new references.

Results We found 22 retracted paper mill articles from 8 journals published between June 21, 2016, and June 16, 2022. Both comparison groups (50 articles each) were restricted to articles published between January 1, 2016, and October 31, 2024. Articles in all 3 groups included similar numbers of total references per article (median [IQR]: 42 [36-46] retracted articles; 47 [36-55] articles in same journals; 51 [45-60] articles in independent journals) and percentages of discussion sentences per article (median [IQR]: 34 [30-38] retracted articles; 40 [30-48] same journals; 34 [26-43] independent journals). Analyses of discussion sections indicated that retracted paper mill articles included higher percentages of references that were first cited in the discussion and higher percentages of discussion sentences that described background information and cited new references (**Figure 25-1022**).

Conclusions Our analyses suggest that some discussion sections in retracted paper mill cancer research articles reiterate background information that is supported by new and possibly superfluous references. While recognizing that genuine studies also cite new references in discussion sections, superficial and redundant second introductions in discussion sections could help readers identify potentially problematic articles in high IF cancer journals, particularly when combined with other features of paper mill support.[1]

References

1. Byrne JA, Abalkina A, Akinduro-Aje O, et al. A call for research to address the threat of paper mills. *PLoS Biol*. 2024;22(11):e3002931. doi:10.1371/journal.pbio.3002931

2. Kesmodel US. Cross-sectional studies—what are they good for? *Acta Obstet Gynecol Scand*. 2018;97(4):388-393. doi:10.1111/aogs.13331

3. Toronto CE, Remington R. Discussion and conclusion. In: *A Step-by-Step Guide to Conducting an Integrative Review*. Toronto CE, Remington R, eds. Springer International; 2020:71-84.

[1]School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia, jennifer.byrne@health.nsw.gov.au; [2]School of Public Health and Social Work, Queensland University of Technology, Kelvin Grove, Queensland, Australia; [3]NSW Health Statewide Biobank, NSW Health Pathology, Camperdown, New South Wales, Australia.

Conflict of Interest Disclosures None reported.

Funding/Support Adrian G. Barnett and Jennifer A. Byrne acknowledge grant funding from the National Health and Medical Research Council of Australia (ideas grant APP2029249). This grant supports Annie Whamond's PhD candidature.

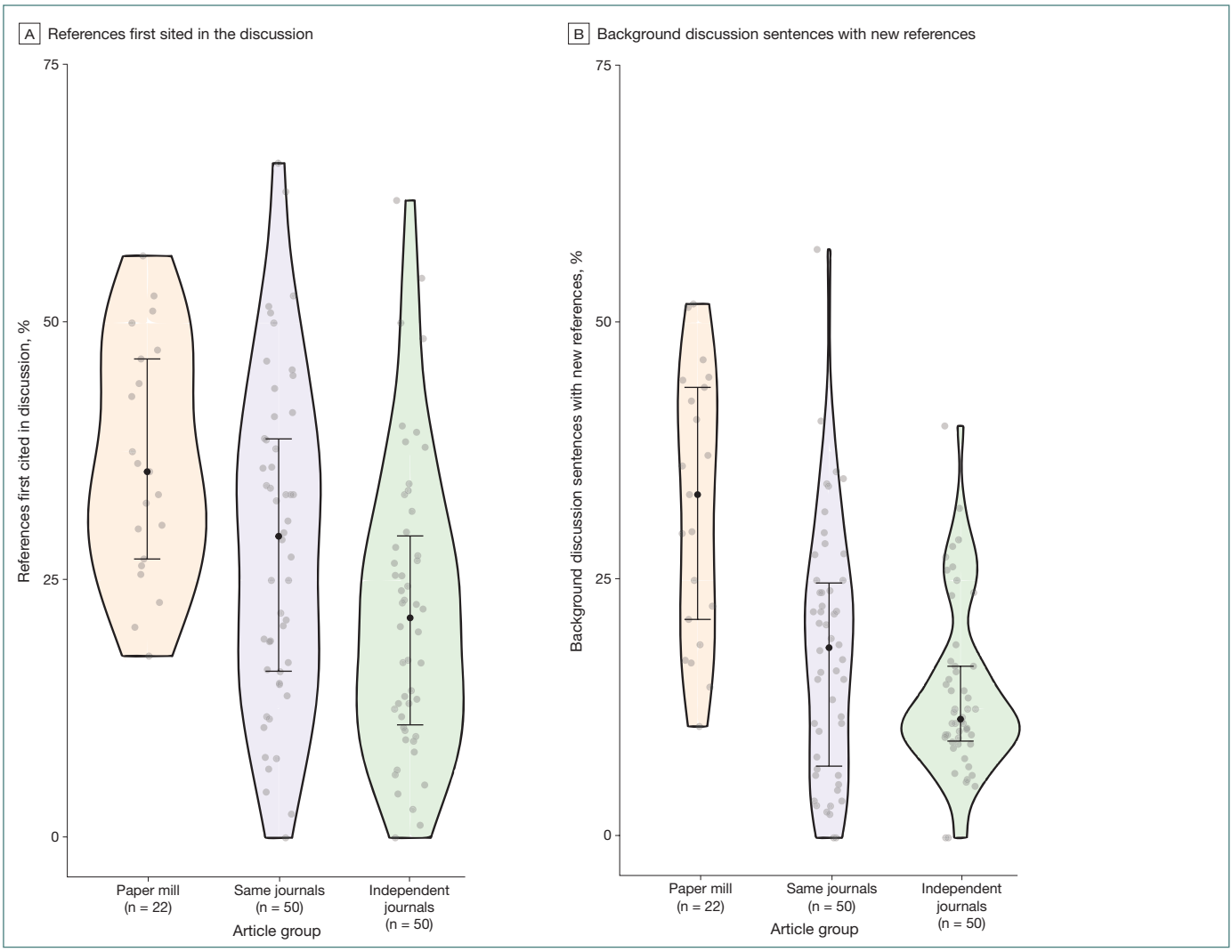
Role of Funder/Sponsor The funding body played no role in the study design, data collection, management, analysis, or interpretation, and will play no role in the writing of any report, or the decision to submit the report for publication.

Screening Articles for Tortured Phrases With a Regular Expressions–Based Detector

Alexandre Clausse,[1] Guillaume Cabanac,[1,2] Pascal Cuxac,[3] Cyril Labbé[4]

Objective The Problematic Paper Screener (PPS) features 10 detectors to identify problematic articles in the scientific record.[1] The detectors search through metadata, references,

Figure 25-1022. Violin Plots of References Cited for the First Time in Discussion Sections and Background Sentences in Discussions Supported by New References in Research Articles That Were Retracted Due to Paper Mill Involvement, Published in the Same Journals as Retracted Articles, or Published in Independent Journals



Dots represent median values, and whiskers indicate IQRs.

or textual contents. The tortured phrases[2] detector uses a list of more than 7000 identified suspect expressions (**Table 25-1159**) to query a database (Dimensions; Digital Science) with approximately 130 million scientific articles.[3] The PPS displays the screening results on a public website. In this qualitative study, we introduce an alternative screening algorithm based on regular expressions (regex) and benchmark its effectiveness. Publishers are welcome to use this stand-alone algorithm that does not require a subscription to the database that we used.

Design As of May 2025, we designed an algorithm that uses regex based on the PPS fingerprints list in the database to capture different forms of tortured phrases, such as "man-made consciousness," "profound learning AND deep learning," and "128 pieces"~5' [128 bits] (ie, a tortured phrase with components in a 5-word sliding window). This approach is independent from search engines, as it matches fingerprints against the textual content of each article, ignoring their metadata, figure and table contents (including captions), and references. We benchmarked this new approach on the Hindawi extensible markup language (XML) corpus, which contains several hundred articles with tortured phrases, focusing on articles published from 2020 to 2022—the period with the most PPS-flagged articles (n = 3400).

Results The regex-based algorithm flagged 2455 problematic articles, with 1948 also flagged by PPS, yielding a 58%

Table 25-1159. Top 5 Tortured Phrases From Retracted Hindawi Articles, as Detected by the PPS in May 2025[a]

Fingerprint tortured phrase	Expected text	No. of retrieved articles
"Information mining"	Data mining	97
"Grouping methods"	Classification methods	48
"Huge information"	Big data	30
"Crown epidemic"	Corona epidemic	26
"Data technology (IT)"	Information technology	26

Abbreviation: PPS, Problematic Paper Screener.

[a]Note that the tortured phrase detector flags articles containing 5 or more tortured phrases.

overlap. After removing duplicates, 401 articles flagged by the PPS were missing from the Hindawi XML corpus, amounting to 3371 problematic articles in 139 different journals. The evaluation is ongoing; we analyzed the top 200 results and found 48 false-positive results due to database indexing issues related to special characters and figures and PPS querying issues. We also found 52 false-negative results due to fingerprints list issues and because we excluded figures, tables, and references. Overall, we obtained 100 true-positive results, as both detectors (ie, PPS and regex-based algorithm) extracted expressions on the fingerprints list from the same articles.

Conclusions The regex-based algorithm yielded results comparable to the current PPS screening process. Other configurations should be tested, including figures, tables, and references in the screening, as tortured phrases can also appear there. We invite publishers to use this regex-based approach to screen incoming manuscripts. We noticed several querying issues with the database that we used, and some fingerprints need to be redesigned. We will investigate further the validation assessment, as there are still 3351 articles to be reassessed. This research could contribute to improving the regex and updating the PPS fingerprints list, while offering detailed feedback and bug reports to the database developer.

References

1. Cabanac G, Labbé C, Magazinov A. The 'Problematic Paper Screener' automatically selects suspect publications for post-publication (re)assessment. *arXiv*. Preprint posted online October 7, 2022. doi:10.48550/arXiv.2210.04895

2. Cabanac G, Labbé C, Magazinov A. Tortured phrases: A dubious writing style emerging in science—evidence of critical issues affecting established journals. *arXiv*. Preprint posted online July 12, 2021. doi:10.48550/arXiv.2107.06751

3. Herzog C, Hook D, Konkiel S. Bringing down barriers between scientometricians and data. *Quantitative Science Studies*. 2020;1:387-395. doi:10.1162/qss_a_00020

[1]Université de Toulouse, IRIT (UMR 5505), Toulouse, France, alexandre.clausse@irit.fr; [2]Institut Universitaire de France, Paris, France; [3]INIST CNRS, UAR 67, Vandoeuvre-lès-Nancy, France; [4]Université Grenoble Alpes, CNRS, Grenoble INP, LIG UMR 5217, Grenoble, France.

Conflicts of Interest Disclosures Guillaume Cabanac is the administrator of the Problematic Paper Screener, a public platform that uses metadata from Digital Science and PubPeer via no-cost agreements. Guillaume Cabanac and Cyril Labbé have been in touch with most of the major publishers and their integrity officers, offering pro bono consulting regarding detection tools to various actors in the field, including Clear Skies, Morressier, River Valley, Signals, and STM.

Funding/Support Alexandre Clausse, Guillaume Cabanac, and Cyril Labbé received funding from the European Research Council. Guillaume Cabanac received funding from the Institut Universitaire de France.

Peer Review

In-person

Librarian and Information Specialist Perceptions of Peer Reviewing Systematic Reviews

Melissa L. Rethlefsen,[1] Carrie Price,[2] Sara Schroter[3]

Objective To explore the perspectives of librarians and information specialists (LISs) who participated in a randomized controlled trial of the effect of LIS involvement on reporting quality of systematic reviews (SRs).[1]

Design We surveyed LISs who completed a peer review of an SR in a trial conducted in *BMJ*, *BMJ Open*, and *BMJ Medicine* from January 3, 2023, to January 2, 2024. LISs were not told they were peer reviewing manuscripts as part of a trial but were sent invitations to review as part of the usual process. The questionnaire sought to understand their experience, what aspects of manuscripts they focused on, perceived impact on editorial decision-making and authors' revisions, and willingness to peer review again. To better understand factors that might impact decisions to review again, we contacted 27 respondents to participate in a focus group concentrating on facilitators and barriers to peer reviewing SRs.

Results Of the 88 LISs invited to participate in the survey, 70 (79.5%) responded. Most respondents had 6 or more years of experience as an LIS (67/70; 95.7%) and advising researchers on doing SRs (55/70; 78.6%) and had peer reviewed for a journal prior to the study (57/70; 81.4%). Most focused on the search and SR methods when reviewing but also commented on aspects such as research question formulation, plagiarism, and study results and conclusions. Two-thirds (44/66; 66.7%) believed they impacted editors' decision-making and 59.1% (39/66) believed they impacted the authors' revisions. Only 3 factors were considered extremely or very likely to impact their decision to review again: their schedule and/or lack of time, review turnaround time, and their sense of professional duty (**Table 24-0830**). Seventeen of 27 invited LISs (63.0%) participated in a focus group. Time was the primary barrier identified in the focus groups, followed by a sense of intimidation. LISs reported being motivated by feeling valued by editors, the enjoyment of peer reviewing, the desire to improve SR quality, and peer review as a learning experience. Several expressed surprise and delight at being asked to peer review for the journals.

Conclusions A select sample of highly engaged LIS respondents believed they made a difference through their peer reviews and said they were very likely to agree to peer review in the future. LISs may be an underused peer reviewing resource with methodological experience that can help editors make decisions and improve the quality of SRs.

Reference

1. Rethlefsen ML, Schroter S, Bouter LM, et al. Improving peer review of systematic reviews and related review types by involving librarians and information specialists as

Table 24-0830. Factors Impacting LISs' Decisions to Peer Review

Factor	Total No. of LISs	Impact on decision, No. (% [95% CI])		
		Extremely/very important	Moderately/ slightly important	Not at all important
Schedule and/or lack of time	66	55 (83 [73-90])	10 (15 [8-26])	1 (2 [0.3-8])
Sense of professional duty	66	42 (84 [52-74])	24 (36 [26-48])	0 (0 [0-0.6])
Review turnaround time	66	38 (58 [46-69])	28 (42 [31-54])	0 (0 [0-0.6])
Opportunity to learn something new	66	26 (39 [28-51])	30 (45 [34-57])	10 (15 [8-26])
Acknowledgment (eg, via ORCID, Web of Science)	66	21 (32 [22-44])	36 (55 [43-66])	9 (14 [7-24])
Contribution of paper to subject area	65	19 (29 [20-41])	39 (60 [48-71])	7 (11 [5-21])
Manuscript topic	66	16 (24 [16-36])	41 (62 [50-73])	9 (14 [7-24])
Desire to keep up to date on current research	66	15 (23 [14-34])	39 (59 [47-70])	12 (18 [11-29])
Acknowledgment from institution or supervisor (eg, for tenure, promotion, and/or scholarship)	66	13 (20 [12-31])	37 (56 [44-67])	16 (24 [16-36])
Relevance of topic to own interests	66	13 (20 [12-31])	42 (64 [52-74])	11 (17 [10-27])
Training	66	10 (15 [8-26])	41 (62 [50-73])	15 (23 [14-34])
Prestige of journal	66	4 (6 [2-15])	43 (65 [53-76])	19 (29 [0-0.6])
Offer of compensation	66	3 (5 [2-13])	27 (41 [30-53])	36 (55 [43-66])
Reputation of the authors of the paper	66	2 (3 [1-10])	15 (23 [14-34])	49 (74 [63-83])

Abbreviations: LIS, librarian and information specialist; ORCID, Open Researcher and Contributor Identifier.

methodological peer reviewers: a randomised controlled trial. *BMJ Evid Based Med*. Published online March 11, 2025. doi:10.1136/bmjebm-2024-113527

[1]Health Sciences Library & Informatics Center, University of New Mexico, Albuquerque, NM, US, mlrethlefsen@gmail.com; [2]ToxStrategies, A Blue Ridge Life Sciences Company, Baltimore, MD, US; [3]BMJ Publishing Group, London, UK.

Conflict of Interest Disclosures This study follows up on a study that was conducted as part of Melissa L. Rethlefsen's self-funded PhD project registered at Maastricht University, the Netherlands, in collaboration with BMJ Publishing Group. Carrie Price was employed at the National Institutes of Health Library at the time of this research. Sara Schroter is a full-time employee of BMJ Publishing Group but is not involved in decision-making on individual research submissions.

An Agent-Based Modeling Approach for Evaluating Interventions to Optimize Peer Review

Abdelghani Maddi,[1] Ahmad Yaman Abdin,[2,3] Francesco De Pretis[4,5]

Objective To evaluate, using an empirically calibrated, agent-based model (ABM), whether a structured reviewer-training intervention is associated with improved peer review quality and reviewer engagement and reduced positive outcome bias across single- and multijournal settings. Existing randomized and quasi-experimental trials of peer review interventions are limited in scope and effect size,[1] and evolutionary models suggest competition among journals shapes reviewer effort.[2]

Design We built an ABM populated by authors, reviewers, and editors. Calibration used the PeerRead International Conference on Learning Representations (ICLR) 2017-2019 corpus (8151 reviews, 4905 submissions).[3] Simulations followed the STRESS-ABM recommendations. Single-journal scenarios assumed 2500 submissions per year, while multijournal portfolios comprised 5 titles handling 10,000 combined submissions. The intervention increased the share of reviewers completing a certified online training module from 0% to 40%. Main outcomes were (1) review quality score (1-10 ordinal), (2) reviewer engagement (report length, number of words), and (3) positive outcome bias (difference in acceptance probability between high- and low-significance manuscripts). Each scenario used 10,000 Monte-Carlo replications; 95% confidence intervals were obtained with percentile bootstrap.

Results In the single-journal baseline, mean review quality score was 5.52 (95% CI, 5.49-5.55), and mean reviewer engagement was 604 words (95% CI, 598-610); positive outcome bias was 0.33 (95% CI, 0.32-0.35) (**Table 25-0961**). With training, mean scores increased to 5.80 (95% CI, 5.77-5.83; Δ +0.28), engagement to 682 words (95% CI, 675-689; Δ +78 words), and bias decreased to 0.26 (95% CI, 0.25-0.28; Δ -0.07). Multijournal results were similar (score, 5.54 vs 5.82; engagement, 607 vs 685 words; bias, 0.34 vs 0.27). Across 10,000 replications per arm, each difference was significant at $P < .001$.

Table 25-0961. Simulated Peer-Review Outcomes With and Without Reviewer-Training Intervention

Outcome	Mean (95% CI)	
	Baseline (n = 10,000)	Training (n = 10,000)
Review quality score (1-10)	5.52 (5.49-5.55)	5.80 (5.77-5.83)
Reviewer engagement (No. of words)	604 (598-610)	682 (675-689)
Positive outcome bias (Δ acceptance)	0.33 (0.32-0.35)	0.26 (0.25-0.28)

Conclusions A pragmatic reviewer training package was associated with modest but consistent improvements in review quality, reviewer engagement, and reduction of positive outcome bias in both single- and multijournal simulations. These data support further real-world trials of low-cost training interventions.

References

1. Heim A, Ravaud P, Baron G, Boutron I. Designs of trials assessing interventions to improve the peer review process: a vignette-based survey. *BMC Med*. 2018;16(1):191. doi:10.1186/s12916-018-1167-7

2. Radzvilas M, De Pretis F, Peden W, Tortoli D, Osimani B. Incentives for research effort: an evolutionary model of publication markets with double-blind and open review. *Comput Econ*. 2023;61(4):1433-1476. doi:10.1007/s10614-022-10250-w

3. Kang D, Ammar W, Dalvi B, et al. A dataset of peer reviews (PeerRead): collection, insights, and NLP applications. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (HLT)*. Association for Computational Linguistics; 2018;1. doi:10.18653/v1/N18-1149

[1]Sorbonne Université, CNRS, Groupe d'Étude des Méthodes de l'Analyse Sociologique de la Sorbonne, GEMASS, Paris, France; [2]Division of Pharmasophy, School of Pharmacy, Saarland University, Saarbrücken, Germany; [3]Division of Bioorganic Chemistry, School of Pharmacy, Saarland University, Saarbrücken, Germany; [4]Department of Environmental and Occupational Health, School of Public Health, Indiana University Bloomington, Bloomington, IN, US, francesco.depretis@unimore.it; [5]Department of Communication and Economics, University of Modena and Reggio Emilia, Reggio Emilia, Italy.

Conflict of Interest Disclosures None reported.

Funding/Support This research was partially supported by the French Research Agency ANR via the OPENIT project (projet AAPG 2024, Agence Nationale de la Recherche).

Role of the Funder/Sponsor The funding organizations 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 GPT-o3 (OpenAI) was used for code and writing and editing on January 31, 2025. All authors take responsibility for the integrity of the content.

Evidence of Use of Template-Based Peer Review Reports and Concern About Review Mills

Cyril Labbé,[1] Gilles Hubert,[2] Wendeline Swart,[2] Guillaume Cabanac[2,3]

Objective Paper mills are well documented, and the existence of review mills has been suggested.[1] Our aim was to test tools and methods that would report evidence of such mills. We studied 4 datasets of peer review reports and found evidence for template-based peer review report practices.

Design Attempting to maximize diversity, we collected 4 datasets gathering 148,159 peer review reports. We used webscraped peer review reports from MDPI journals (11 journals from 2018-2025: 47,593 articles and 122,831 reports), *BMJ* (97 articles from 2021-2024: 308 reports), *PeerJ* journals listed in the Multidisciplinary Open Peer Review Dataset (MOPRD)[2] (7 journals from 2015-2022: 6292 articles and 12,959 reports), as well as the available dataset NLPeer[3] (12,061 reports from 5 conferences in NLP: years 2016, 2017, 2020, and 2022, and F1000Research platform: year 2022). We investigated only plain text reports (excluding attached files) considering "round 1" only. We computed statistics on report length, common sequences of terms (CST), and similarity measures between reports.

Results Depending on the dataset, the mean report length ranged from 251 to 530 words, and the median from 197 to 433 words. Reports with fewer than 20 words (from 1 to 2825, depending on the dataset) were mostly coming from resubmissions inaccurately reported as a round 1 submission. Regarding CST, excluding stop words (respectively including stop words), 8% to 12% (respectively 11.16% to 24.72%) of reports shared common sequences of at least 10 terms. This represented from 15% to 28% (CST excluding stop words) of articles having at least 1 such report. Regarding similarity, approximately 0.5% to 1.3% of reports were highly similar to another report (ie, cosine similarity excluding stop words greater than 0.75). This represented 0.8% to 3% of articles having at least 1 such report. After having automatically highlighted common CST in reports sharing more than 100 words, we analyzed them qualitatively in order to identify potential templates used by reviewers to write their reports. We were able to identify very generic chunks that are reused from report to report, sometimes by different identifiable reviewers.

Conclusions Our results showed evidence of template-based peer review reports practices. We observed that the extent of these practices varies from one dataset to another. This might be due to reviewers' practices, discrepancies in dataset sizes, differences in scientific fields, or to journals', editors', or publishers' habits. More openly available review report datasets would help further characterize and understand this phenomenon.

References

1. Oviedo-García MÁ. The review mills, not just (self-) plagiarism in review reports, but a step further. *Scientometrics*. 2024;129:5805-5813. doi:10.1007/s11192-024-05125-w

2. Lin J, Song J, Zhou Z, Chen Y, Shi X. MOPRD: a multidisciplinary open peer review dataset. Neural Comput Applications. 2023;35:24191-24206. https://doi.org/10.1007/s00521-023-08891-5

3. Dycke N, Kuznetsov I, Gurevych I. NLPeer: a unified resource for the computational study of peer review. *arXiv*. Preprint posted online November 12, 2022.doi:10.18653/v1/2023.acl-long.277

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG (Laboratoire d'Informatique de Grenoble LIG - équipe SIGMA), Domaine Universitaire de Saint-Martin-d'Hères, Grenoble, France, cyril.labbe@imag.fr; [2]IRIT, Université de Toulouse, CNRS, Toulouse INP, UT, Laboratoire IRIT - équipe IRIS, Toulouse, France; [3]Institut Universitaire de France (IUF), Paris, France.

Conflict of Interest Disclosures None reported.

Funding/Support Cyril Labbé and Guillaume Cabanac acknowledge the NanoBubbles project (https://nanobubbles.hypotheses.org/) that has received Synergy grant funding from the European Research Council (ERC) within the European Union's Horizon 2020 program, grant agreement number 951393. Guillaume Cabanac received funding from the Institut Universitaire de France (IUF).

Additional Information Cyril Labbé and Gilles Hubert (gilles.hubert@irit.fr) are co–corresponding authors.

Virtual

Research Culture Influences in Peer Review: A Targeted Thematic Analysis of Current Challenges in Peer Review

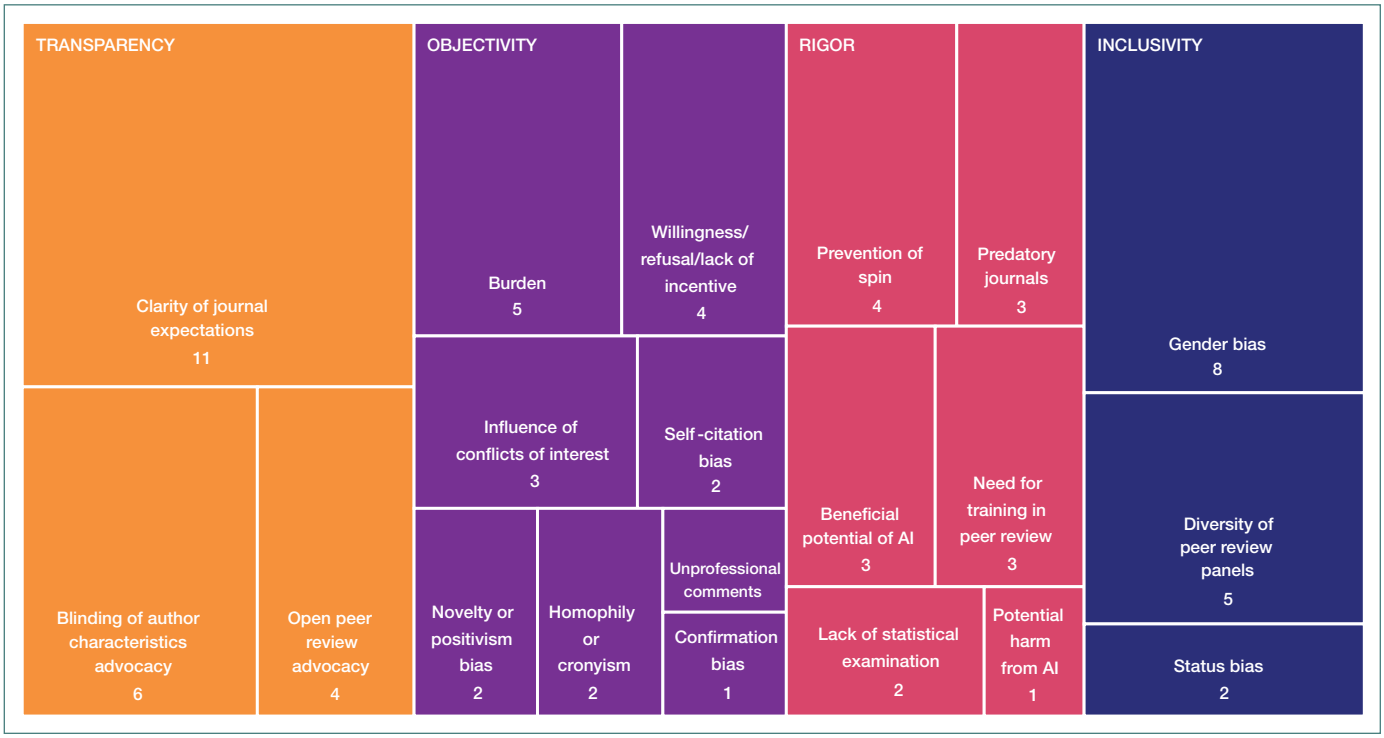
Lesley Uttley,[1] Louise Falzon[1]

Objective This study aimed to identify current challenges in peer review relating to research culture using a targeted thematic analysis of research culture in health and biomedical research from a wider scoping review.[1]

Design To understand challenges of peer review influenced by research culture, the literature searches of a wider scoping review[1] were subject to a focused qualitative thematic analysis relating specifically to peer review, following PRISMA-ScR (scoping review) guidelines[2] and adhering to Nowell et al's[3] trustworthiness criteria for thematic analysis. PubMed and Web of Science were searched for empirical studies published between January 2012 and April 2024 in health and biomedical research culture. Eligible articles for this study were those that reported relevance to peer review (regarding manuscript publication or funding applications) in the title or abstract. Editorials, discussion pieces without data, and articles relating to peer review for conferences were excluded. The main outcome was the qualitative themes from the included articles that were associated with peer review. Relevant papers were subjected to inductive content analyses by 2 authors (L.U. and L.F.), which involved noting key findings to identify recurring themes relating to the research culture topic or finding discussed. The themes were then deductively categorized according to a framework for research integrity across 4 domains: inclusivity, transparency, rigor, and objectivity.[1]

Results Fifty-three papers were included from a total of 1680 citations generated by literature searches. Relevant themes (n = 20) regarding the challenges of peer review, ordered by prevalence, were reviewed for thematic content across the 4 domains (**Figure 25-0850**). Inclusivity included gender biases (n = 8), status bias (n = 2), and diversity of peer review panels (n = 5); transparency included clarity of journal expectations (n = 11), open peer review advocacy (n = 4), and blinding of author characteristics (n = 6); rigor included predatory journals (n = 3), prevention of spin (n = 4), lack of statistical examination (n = 2), potential benefits of artificial intelligence (n = 3), need for training in peer review (n = 3), and potential harms from artificial intelligence (n = 1); and objectivity included willingness/refusal/lack of incentive (n = 4), burden (n = 5), influence of conflicts of interest of reviewers (n = 3), unprofessional comments (n = 1), self-citation bias (n = 2), novelty or

Figure 25-0850. Treemap of Peer Review Themes Across 4 Domains of Research Integrity



positivism bias (n = 2), confirmation bias against innovative research (n = 1), and homophily (preference for shared characteristics of authors) or cronyism (n = 2).

Conclusions A variety of research culture influences underpin peer review challenges in the traditional academic publishing model. Themes pertain to 4 cornerstones of research integrity[1] relating to inclusive workforces, transparency of processes, use of best practice methods, and the objectivity (and incentives) of peer review. Emerging threats from technological innovations in conjunction with incentives fuelled by research culture are likely shaping the research landscape. Sustained and joined-up conversations between authors, publishers, and reviewers about potential peer review innovations are required to safeguard genuine research and aid the dissemination of findings that promise authentic advancements in scientific research.

References

1. Uttley L, Falzon L, Byrne JA, et al. Research culture influences in health and biomedical research: rapid scoping review and content analysis. *J Clin Epidemiol*. 2024;178:111616. doi:10.1016/j.jclinepi.2024.111616

2. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467-473. doi:10.7326/M18-0850

3. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. *Int J Qual Methods*. 201728;16(1). doi:10.1177/1609406917733847

[1]School of Medicine and Population Health, University of Sheffield, UK, l.uttley@sheffield.ac.uk.

Conflict of Interest Disclosures None reported.

Funding/Support Lesley Uttley is supported by a grant from the UKRI Medical Research Council (MR/Z504063/1).

Motivations to Participate in the Peer Review Process at the *Journal of Urology*

Anne G. Dudley,[1] George Koch,[2] Kyle Rose,[3] Roei Golan,[4] Jennifer Regala,[5] Casey Seideman,[6] Amanda North,[7] Kevin Koo,[8] Kevan Sternberg,[9] Gina Badalato,[10] Benjamin Dropkin,[11] Nicholas Chakiryan,[12] Robert Siemens,[13] Peter Clark,[14] Andrew Harris[11]

Objective Peer review is a critical aspect of academic publishing, yet the process takes significant time and energy for the reviewer and is a voluntary activity. Current surveys report high levels of urologist burnout, and recent events, including the COVID-19 pandemic, have led to a shift toward personal priorities outside of work potentially limiting reviewer pools. Within urology, editors report difficulty finding appropriate numbers of peer reviewers for submitted manuscripts. We sought to assess motivations to participate in the peer review process within a pool of recent *Journal of Urology* reviewers.

Design *The Journal of Urology* partnered with members of the American Urologic Association publications team to develop and administer a web-based survey to a diverse group of reviewers from September 1 to December 31, 2023. All authors and reviewers over the preceding 3 years were invited to participate. The survey addressed various aspects including career stage, their experience as reviewers, and peer review process challenges, incentives, motivators, and feedback needs.

Results Respondents (n = 275) completed an average of 9 reviews in the past 12 months and reported 16 years of experience as reviewers. Most reviewers were experienced urologists less than 11 years from training (64% [176]) with only 7% (18) currently in training (resident/fellow). Time emerged as a key variable with 86% (236) of respondents declining additional reviews due to time constraints. A total of 67% (184) of respondents reported reviewing time was worthwhile, yet only 35% (96) felt appropriately recognized for time and effort, and 55% (151) reported incentives would increase time spent on a peer review. Motivations to review included "to give back" (80% [220]), "to learn" (71% [195]), "to get involved" (61% [168]), and "to grow my career" (39% [107]). Most respondents (91% [250]) read other reviewers' reviews to learn. When asked to select specific incentives to review more papers, American Urologic Association products such as waived meeting fees and membership were highly valued (64% [176]; 62% [170]), followed by recognition by local department leadership (43% [118]) and money (40% [109]). Only 14% (38) of respondents desired gear or swag, and only 23% (63) desired to be named in the journal alongside the manuscript.

Conclusions Peer review motivations are diverse and suggest that urologists participate for professional development and an ongoing desire to learn and participate in the field as a whole. Study limitations include nonresponder bias, limited survey period, and lack of granular data on personal and professional motivators. Time remains an important constraint, but incentives may increase allocated time for academic pursuits. Professional meeting/membership fee waivers may be motivators to increase participation. Local efforts to recognize reviewers within departments may work synergistically to increase available reviewers and fulfill career development goals.

[1]Connecticut Children's, Hartford, CT, US, annedudleymd@gmail.com; [2]The Ohio State University Wexner Medical Center, Columbus, OH, US; [3]Ochsner Medical Center, New Orleans, LA, US; [4]Florida State University School of Medicine, Gainesville, FL, US; [5]Wolters Kluwer Health, Baltimore, MD, US; [6]Doernbecher Children's Hospital at OHSU, Portland, OR, US; [7]The Children's Hospital at Montefiore, Bronx, NY, US; [8]Mayo Clinic College of Medicine and Science, Rochester, MN, US; [9]Northwestern Medical Center, Chicago, IL, US; [10]Columbia University, New York, NY, US; [11]University of Kentucky, Lexington, KY, US; [12]H Lee Moffitt Cancer Center, Tampa, FL, US; [13]Queen's University, Kingston, ON, Canada; [14]Levine Cancer Institute, Charlotte, NC, US.

Conflict of Interest Disclosures None reported.

Acknowledgment We thank Martha Keyes and the *Journal of Urology* publications staff for their assistance with this initiative.

Review of Proposals Submitted to Elsevier's Peer Review Workbench

Bahar Mehmani,[1] Silvia Dobre,[1] Ramadurai Petchiapppan[1]

Objective To encourage evidence-based peer review studies, Elsevier launched the Peer Review Workbench (PRW) in 2022. Here we report on the studies that have since worked in PRW, as we believe peer review studies at large scale are needed for the progress of research on research and understanding how to improve the quality of the peer review process.

Design PRW is a curated and anonymized dataset of all Elsevier journal manuscript and peer review metadata on more than 5 million authors and reviewers. In our white paper,[1] we describe all the attributions and results of some calculations such as variation of interrater agreement, number of peer reviewers per manuscript, peer review rounds, gender balance, and country representation in editorial, authorship, and reviewers' composition across different disciplines and impact factor quartiles. Interested researchers are provided with a proposal protocol[2] to submit their research plan. All protocols go through rounds of peer review by the academic advisors of the PRW, who are listed in our page.[2] The accepted applications receive an onboarding training by our data scientist (R.P.) who further prepares the relevant data segment suitable for their primary and secondary research purposes.

Results Since September 2022, PRW has received 12 proposals, of which 6 have been accepted. Of the 6 accepted proposals, all have been onboarded and are actively working with the data. Thirty-one researchers from 17 universities in 9 countries are involved in these research projects. Based on the self-reported data, 39% (n = 12) of these researchers are women and 61% (n = 19) are men; the options "Other" and "Prefer not to disclose" were both at 0%. We promote transparency in science and therefore publish all accepted proposals on a dedicated Peer Review Workbench Research Paper Series page on SSRN.[2] **Box 25-0963** shows the list of research proposals currently working with the data. So far, these proposals published on SSRN have attracted over 170 downloads and almost 800 views.

Conclusions The breadth of the ongoing studies encourages us to continue adding Elsevier journal and manuscript metadata annually to PRW and organize our seasonal seminars starting from autumn 2025 to report on the progress and findings of each study. This may also help improve the visibility of PRW.

References

1. Petchiappan R, James K, Plume A, et al. Analysing Elsevier Journal Metadata within peer review Workbench. SSRN. Revised June 9, 2025. Accessed July 16, 2025. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4211833

Box 25-0963. List of Active Studies on Peer Review Workbench

Quantifying strictness of peer evaluation of peer review research

Roles of different countries in science gatekeeping

The role of disagreement in scientific conservatism

Measuring the effect of COVID-19 on the reviewing process

Is the resilience of the academic peer-review system breaking down?

Reviewer reliability and the extent of randomness in publication decisions

2. SSRN. PRW: Peer Review Workbench Research Paper Series. Accessed July 16, 2025. https://www.ssrn.com/index.cfm/en/peer-review-workbench/

[1]Elsevier, Amsterdam, the Netherlands, bahar.mehmani@elsevier.com.

Conflict of Interest Disclosures All authors are employees of Elsevier, the owner of the Peer Review Workbench, the publisher of the journals whose metadata is hosted on PRW, and the owner of the submission system used for collecting the data. Elsevier is also the owner of the Scopus database that was used to enrich the journals' metadata with their different impact factor quartiles and subject areas. Bahar Mehmani is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Peer Review Process and Models

In-person

Efficiency of Author Anonymization in Peer Review

Markus K Heinemann,[1] Andreas Boening,[2] Kazunori Okabe,[3] Jessica Bogensberger,[4] Zulfugar Timur Taghiyev[2]

Objective Double-anonymized peer review is thought to enhance objectivity but causes additional work for editorial staff.[1-3] The aim of this study was to evaluate its efficiency in a midsized cardiothoracic surgical journal.

Design In July 2017, the journal introduced anonymization, performed by a secretary during the first editorial office check, and introduced custom questions to the review form asking reviewers if they had an idea about the origin of a manuscript despite its anonymization and, if so, why. Descriptive statistics were used for analysis, and a 2-proportion *z* test was utilized as applicable.

Results Between July 2017 and November 2024, a total of 1735 manuscripts were amenable for anonymization and sent for peer review to a mean of 2 reviewers. For 525 manuscripts (30.3%), at least 1 reviewer had an idea about its origin for at least 1 of 4 potential reasons (multiple entries possible; n = 597): faulty anonymization (232), references (172), familiar with work (146), and characteristic language (47). In 82

manuscripts (15.6%), more than 1 reason was given by at least 1 reviewer. Forty guesses (7.6%) were incorrect for the following reasons: misled by impressions of being familiar with the work (20), references (9), faulty anonymization (6), and characteristic language (5). Multiple reviewers guessed for the same reason in 8 cases and for different reasons in 32 cases. One of the reviewers was incorrect in 5 manuscripts; all reviewers guessed correctly in 35. Among 649 accepted manuscripts (37.4% of the total manuscripts), 223 (34.3%) were deanonymized by reviewers, whereas 426 stayed anonymized (65.6%). The 2-proportion *z* test demonstrated higher acceptance rates for anonymized compared with deanonymized manuscripts (426 vs 223; $z = -11.27$).

Conclusions Perfect anonymization remains challenging, if not impossible, because hints identifying origins can be hidden throughout a manuscript. Roughly 30% of manuscripts were correctly deanonymized by the reviewers of an admittedly limited specialty surgical community. Even though the results seem to support different editorial dispositions of anonymized vs deanonymized manuscripts, editors must be aware of the confined reliability of anonymization and weigh it against the additional administrative workload.

References

1. Parmanne P, Laajava J, Järvinen N, Harju T, Marttunen M, Saloheimo P. Peer reviewers' willingness to review, their recommendations and quality of reviews after the Finnish Medical Journal switched from single-blind to double-blind peer review. Res Integr Peer Rev. 2023;8:14. doi:10.1186/s41073-023-00140-6

2. Ucci MA, D'Antonio F, Berghella V. Double- vs single-blind peer review effect on acceptance rates: a systematic review and meta-analysis of randomized trials. *Am J Obstet Gynecol MFM*. 2022;4(4):100645. doi:10.1016/j.ajogmf.2022.100645

3. Kmietowicz, Z. Double blind peer reviews are fairer and more objective, say academics. *BMJ*. 2008;336(7638):241. doi:10.1136/bmj.39476.357280.DB

[1]German Society for Thoracic and Cardiovascular Surgery (DGTHG), Germany, heinemann@uni-mainz.de; [2]Universitaetsklinik Giessen, Germany; [3]Bell Land General Hospital, Osaka, Japan; [4]Thieme Publishers, Stuttgart, Germany.

Conflict of Interest Disclosures Jessica Bogensberger is an employee of Thieme Publishers, the publishing house of the journal investigated.

Manuscript Submissions Following Implementation of Guaranteed Peer Review

Yurong Fei-Bloom,[1] Matthew Welch[2]

Objective *Molecular Biology of the Cell* (*MBoC*) serves as our community cell science journal for members of the American Society for Cell Biology (ASCB) and cell scientists worldwide. ASCB members contribute high-quality research manuscripts to MBoC, enhancing its reputation as a platform for enduring work. Nearly half of the articles published in *MBoC* come from ASCB members, and most submissions from members are accepted. In our annual surveys, ASCB members expressed great value in having a community-led society journal where they can publish their research, which ensures that funds used to publish remain within our community to support meetings, awards, mentorship, and professional training. The objective of this study was to examine the implications of guaranteed peer review for submissions from ASCB members.

Design Starting in January 2024, *MBoC* guaranteed that all newly submitted research manuscripts with current ASCB members as corresponding authors would undergo the peer review process.[1] It is important to note that guaranteed peer review did not imply guaranteed acceptance; editorial standards[2] remained the same for all submissions, whether from ASCB members or non-members. Editors were aware of which manuscripts were submitted by members, but reviewers were not aware of the member status. The Editor in Chief carefully assessed each nonmember submission to determine whether it should be desk rejected or moved forward to the peer-review stage. The peer-review process was overseen by dedicated monitoring editors with subject matter expertise. This structure enabled these editors to manage the peer review effectively, fostering a fair and unbiased system that upholds the integrity of the review for all submissions, including those from members.

Results *MBoC* received 120 submissions (new or revisions) from ASCB members in 2023 and 204 in 2024, which is a 70% increase from the previous year. The total submissions rose by 13%, from 373 in 2023 to 421 in 2024. In 2023, the journal accepted 180 manuscripts, including 102 (57%) from members. In 2024, it accepted 165 manuscripts, with 140 (85%) from members. In 2023, the journal rejected 180 manuscripts, including 12 from members. In 2024, it rejected 233 manuscripts, including 12 from members. The rejection rate of the member manuscripts was 9% in 2023 and 6% in 2024, suggesting that our members consistently submit high-quality work. The average time from submission to the first decision for peer-reviewed articles was 34 days, 2 days fewer than in 2023 (36 days).

Conclusions Findings of this study suggest that guaranteed peer review increased submissions from ASCB members while maintaining the quality and timing of the peer review process. The data also suggest that the community values straightforward and guaranteed access to constructive peer feedback.[3]

References

1. Fei-Bloom Y, Welch M. New ASCB Member Benefit—Guaranteed Peer Review in *MBoC*. ASCB Member News, Publishing, Society News. December 22, 2023. Accessed July 15, 2025. https://www.ascb.org/society-news/new-member-benefit-guaranteed-peer-review-in-mboc/

2. Baum B. Reviewing papers as you would like your papers to be reviewed. *Mol Biol Cell*. 2019;30:3013-3014. doi.org/10.1091/mbc.E19-05-0273

3. Drubin DG. Any jackass can trash a manuscript, but it takes good scholarship to create one (how *MBoC* promotes civil and constructive peer review). *Mol Biol Cell*. 2011;22:525-527. doi.org/10.1091/mbc.e11-01-0002

[1]American Society for Cell Biology, Rockville, MD, US, yfeibloom@ascb.org; [2]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, US.

Conflict of Interest Disclosures None reported.

Editor Initial Manuscript Review: A Masked Pilot Study

Douglas K. Novins,[1] Mary K. Billingsley,[2] Robert R. Althoff[3]

Objective For many journals in science and medicine, a large number of submitted manuscripts may be rejected by editors before peer review. Given the high rates of such "desk rejections," it is important that editorial reviews minimize bias in this decision-making process. One potential source of bias is the editor having access to information regarding the authors of manuscripts, which may result in editorial decision being influenced by authors' academic reputation as well as other author characteristics.[1,2] Unfortunately, a recent review was unable to identify any empirical studies of such "triple-blind" peer review.[3] In this pilot study, we masked author information in the initial stage of editorial review to explore whether this changes patterns in the articles that our journal sends out for peer review.

Design The *Journal of the American Academy of Child and Adolescent Psychiatry* utilizes a 2-round editor in chief (EIC) review process during which the EIC is aware of author identity in both rounds of review. In the standard round 1 review, the EIC completes a high-level reading of the manuscript and the cover letter, checks for test recycling, and reviews study preregistration. In the standard round 2 review, the EIC completes a detailed read of the manuscript. In this preregistered pretest/posttest study (https://doi.org/10.17605/OSF.IO/TNYFX), we followed the above standard process in the pretest. In the posttest, the identity of the authors was masked in the round 1 review and tasks that unmask author identity, such as reading the cover leader, were moved to the round 2 review.

Results A total of 95 manuscripts were included in the pretest and 102 were included in the posttest. In the masked process, significantly more manuscript decisions were made in round 2 (49/102) than in the unmasked process (26/95) ($\chi^2 = 8.91$; 1 degree of freedom; $P < .01$) (**Figure 25-1112**). Overall, EIC decisions (reject, transfer, assignment to an action editor) did not differ between the pretest and posttest ($\chi^2 = 2.48$; 2 degrees of freedom; $P = .29$).

Conclusions The shift of decisions from round 1 to round 2 in the posttest review process is consistent with the shift of key EIC tasks to the round 2 review, resulting in more manuscripts also receiving a detailed read of their manuscript than in the pretest.

References

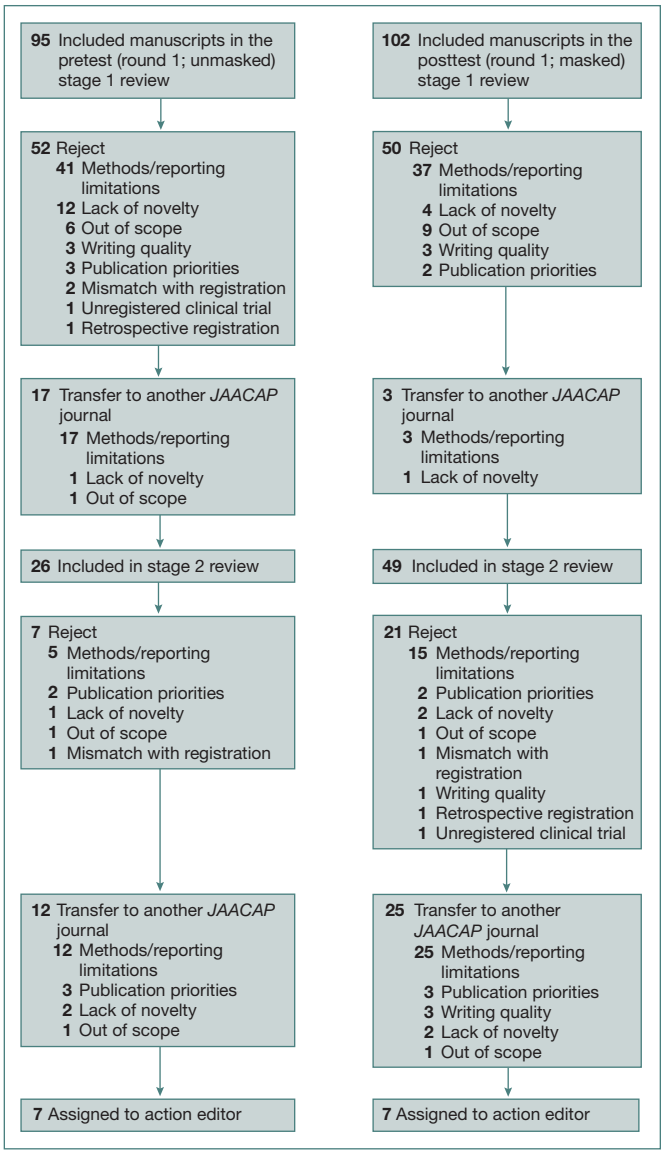
1. Fox CW, Meyer J, Aimé E. Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Funct Ecol*. 2023;37(5):1144-1157. doi:10.1111/1365-2435.14259

2. Srivastava DS, Bernardino J, Marques AT, et al. Editors are biased too: an extension of Fox et al. (2023)'s analysis makes the case for triple-blind review. *Funct Ecol*. 2024;38(2):278-283. doi:10.1111/1365-2435.14483

3. Polnaszek BE, Mei J, Cheng C, et al. Triple-blind peer review in scientific publishing: a systematic review. *Am J Obstet Gynecol MFM*. 2024;6(4):101320. doi:10.1016/j.ajogmf.2024.101320

[1]*Journal of the American Academy of Child and Adolescent Psychiatry* (University of Colorado Anschutz Medical Campus); [2]*Journal of the American Academy of Child and Adolescent Psychiatry* (American Academy of Child and Adolescent

Figure 25-1112. Manuscript Flow During Pretest and Posttest Editor in Chief Review



More than 1 reason could be reported for each decision.

Psychiatry); [3]*Journal of the American Academy of Child and Adolescent Psychiatry* (University of Vermont).

Conflict of Interest Disclosures Douglas K. Novins reported receiving honorarium for serving as editor in chief for the *Journal of the American Academy of Child and Adolescent Psychiatry*, American Academy of Child and Adolescent Psychiatry and grant support from the National Institutes of Health. Robert R. Althoff reported receiving honorarium for serving as associate editor for the *Journal of the American Academy of Child and Adolescent Psychiatry* from the American Academy of Child and Adolescent Psychiatry; grant support from National Institutes of Health and Klingenstein Third Generation Foundation; and stock or Equity in WISER Systems, LLC.

Acknowledgment We acknowledge the contributions of our journal's senior editorial team in designing this study.

Optimizing Proposal Assignments in the Distributed Peer Review System of the World's Largest Radio Telescope Observatory

Andrea Corvillon,[1] John Carpenter,[1] Nihar B. Shah[2]

Objective As the Atacama Large Millimeter/Submillimeter (ALMA) telescope, the largest radio telescope in the world, transitioned from panel-based to distributed peer review to manage increasing proposal volumes, new challenges emerged in aligning reviewer expertise with proposal content. Building on prior work highlighting the importance of reviewer-proposal match quality, this study evaluates a machine learning and optimization framework to improve assignment fairness and accuracy.

Design The new assignment process has 2 steps: (1) using machine learning to measure similarity between a proposal's topic and a reviewer's expertise and (2) optimizing assignments based on a fairness metric. Proposal topics were inferred using latent Dirichlet allocation (LDA), trained on ALMA proposals submitted between 2012 and 2023; the topics were known based on a set of keywords selected by the principal investigators. Reviewer expertise was estimated using the same model trained on each reviewer's past proposals. Both were represented as topic vectors, and cosine similarity was used to assess alignment. Each proposal was assigned to 10 reviewers, and each reviewer received 10 proposals. Assignments were optimized using the PeerReview4All algorithm,[1] which prioritizes proposals with the lowest similarity scores and least reviewer availability, ensuring fairer matches by improving assignments for the most disadvantaged proposals. This method was compared with ALMA's approach from 2021 to 2022, which relied on direct keyword overlap. In 2021 and 2022, 1016 and 1087 reviewers assessed 1497 and 1729 proposals, resulting in 14,970 and 17,290 assignments, respectively.

Results The new method was implemented in 2023, involving 1098 reviewers and 1635 proposals. Similarity scores were validated using survey data from cycles 2021 and 2022. On average, 89% of the reviewers rated their expertise. The median similarity of assignments in which reviewers identified themselves as experts was 0.35 compared with 0.04 for nonexpert assignments, confirming the metric's reliability. Retrospective application of the similarity metric to cycles 2021 and 2022 using the same LDA model showed that the median similarity increased from 0.20 in 2022, the final cycle of the old algorithm, to 0.71 in 2023, the first cycle of the new algorithm, demonstrating a better alignment between reviewers and proposal assignments. Every cycle, on average, 87% of reviewers rated their expertise level on their assigned proposals. With the new algorithm, the percentage of reviewers identifying themselves as experts increased from 45% to 65%, while self-identified nonexperts decreased from 10% to 5%. This tendency continued in cycle 2024, when the new algorithm was also used (**Figure 25-1127**). Additionally, the new algorithm eliminated manual reassignments, reducing manual effort time from 3 to 5 days to nearly zero. These findings will be updated with analysis of new data from 2024 and 2025.

Conclusions ALMA's machine learning–based assignment framework improved reviewer-proposal alignment, increased expertise match rates, and eliminated manual reassignments.

Reference

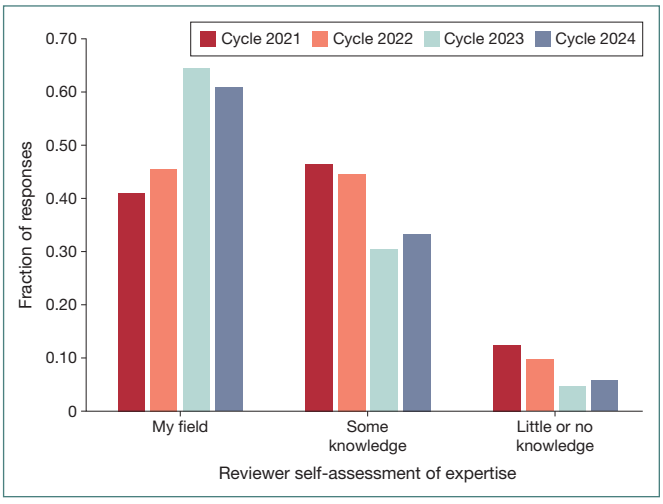
1. Stelmakh I, Shah N, Singh A. PeerReview4All: fair and accurate reviewer assignment in peer review. *J Machine Learning Res*. 2021;22(163):1-66. Accessed July 16, 2025. https://jmlr.csail.mit.edu/papers/volume22/20-190/20-190.pdf

[1]Joint ALMA Observatory, Santiago, Chile, andrea.corvillon@alma.cl; [2]Carnegie Mellon University, Pittsburgh, PA, US.

Conflict of Interest Disclosures Andrea Corvillon and John Carpenter are employed by the Joint ALMA Observatory (JAO), which is jointly managed by Associated Universities, Inc/National Radio Astronomy Observatory, the European Organization for Astronomy Research in the Southern Hemisphere (ESO), and the National Astronomical Observatory of Japan (NAOJ) on behalf of the ALMA partnership. Nihar B. Shah is employed by the Carnegie Mellon University and is a member of the Peer Review Congress Advisory Board but was not involved in the editorial review or decision for this abstract.

Funding The study was funded by grant 1942124 from the National Science Foundation (NSF) to the JAO.

Figure 25-1127. Histogram of Reviewers' Self-Assessment of Their Proposal Assignments in Cycles 2021 to 2024



Role of Funder/Sponsor This work is carried out as part of the duties of the authors at the Joint ALMA Observatory, which receives funding from the NSF through NRAO. The NSF did not contribute to the analysis presented in this contribution.

Acknowledgments ALMA is a partnership of ESO (representing its member states), NSF (US), and the National Institute of Natural Sciences (Japan), together with the National Research Council of Canada (Canada), the National Science and Technology Council and the Institute of Astronomy & Astrophysics, Academia Sinica (Taiwan), and the Korea Astronomy and Space Science Institute (Republic of Korea), in cooperation with the Republic of Chile.

AI-Augmented Peer Review, Collaboration Dynamics, and Human Reviewer Performance

Ashia Livaudais,[1] Dmitri Iourovitski[1]

Objective As AI models improve in quality and affordability, their role in scientific evaluation grows increasingly relevant.[1,2] We investigated the quality of reviews produced by AI tools and humans, alone or in combination (ie, AI augmented), reviewer accuracy at distinguishing AI-generated and AI-augmented peer reviews from human reviews, and whether awareness of AI augmentation affected human reviewers' perceptions of review quality.

Design This mixed-methods study was conducted from July to September 2024 and received institutional ethical approval. We defined peer review subtasks[3] (eg, evaluation of methodological rigor) by analyzing reviews posted to a selection of 50 manuscripts on OpenReview. We selected 100 manuscripts in physics, mathematics, and machine learning and identified 133 participants across 3 continents without conflicts of interest to review the manuscripts and evaluate review quality. The 60 reviewers and 73 meta-reviewers included 48 senior/faculty researchers, 13 journal editors, 42 industry researchers, 23 graduate students, and 7 masters' degree students. Reviewers provided peer review reports for selected manuscripts, and meta-reviewers rated peer review quality using a 0 to 5 scale, with higher scores indicating higher quality. Participants were randomly assigned roles. Manuscripts were anonymized and reviewed by Symby (Symby Labs) (a new AI tool fine-tuned for scientific evaluation) GPT-4 (OpenAI), and Claude Sonnet 3 (Anthropic), alone and in combination with humans (AI-augmented) as well as by humans alone. Half of reviewer participants were informed their reviews might be AI-augmented regardless of actual augmentation. Half of meta-reviewers were informed of potential AI involvement and were asked to also detect AI-generated reviews. Statistical analysis used 1-way ANOVA to compare review quality scores, with post-hoc Tukey Honest Significant Difference tests. Independent sample *t* tests compared review quality scores among informed and uninformed groups. χ^2 testing assessed AI detection accuracy.

Results Review quality scores were highest for AI-augmented human reviews: Symby+ human (4.2), GPT-4 + human (3.9), Symby (3.8), Claude Sonnet 3 + human (3.5), Claude Sonnet 3 (3.4), human (3.3), and GPT-4 (3.1). Human reviewers informed that reviews may be augmented with AI produced review outputs that received higher scores compared with uninformed reviewers (3.6 vs 3.2; $t_{58} = 2.4$, $P = .02$). Informed meta-reviewers gave lower scores overall (3.0 vs 3.7; $t_{71} = -3.1$, $P = .003$). Meta-reviewers from all disciplines had 39% accuracy in distinguishing AI-generated reviews ($\chi^2 = 0.6$; $P = .44$).

Conclusions We found AI-augmented human reviews were ranked higher quality than human-only and AI-only reviews, suggesting AI-augmented human reviews could provide feedback on par with or superseding humans. Awareness of AI-augmentation affected reviewer ratings. Meta-reviewers did not accurately distinguish AI-generated reviews. Study limitations include a convenience sample drawn from 3 quantitative disciplines, review quality measured with a simple scale, no accounting for clustering of reviews within manuscripts, English-language only participation, and participants' awareness of being studied.

References

1. Latona GR, Ribeiro MH, Davidson TR, Veselovsky V, West R. The AI Review Lottery: Widespread AI-Assisted Peer Reviews Boost Paper Scores and Acceptance Rates. *arXiv*. Posted online May 3, 2024. doi:10.48550/arXiv.2405.02150

2. Checco A, Bracciale L, Loreti P, et al. AI-assisted peer review. *Humanit Soc Sci Commun*. 2021;8(25). doi:10.1057/s41599-020-00703-8

3. Bornmann L. Scientific peer review. *Annual Review of Information Science and Technology*. 2013;45(1):197-245. doi:10.1002/aris.2011.1440450112

[1]SymbyLabs, Huntsville, AL, US, livaudais@symbyai.com.

Conflict of Interest Disclosures While the specific version of the Symby tool described in this manuscript is not being commercialized, the authors are involved in ongoing research and development activities that may lead to future commercial applications based on similar technological approaches and methodologies.

Acknowledgment We extend our gratitude to the 133 participants—ranging from master's degree students to senior editors—who contributed their time and expertise to this study. We also acknowledge the OpenReview platform for providing access to review data that informed the design of our study.

Virtual

Use of a 3-Round Modified Delphi Process to Support Rapid Peer Review

Sean Hays,[1] Tyler Carneal,[1] Christopher Kirman[1]

Objective A pilot study was conducted in which a community of SciPinion science experts (>32,000 registered users) and the SciPinion web platform performed rapid peer review of a study.

Design Review design elements were included to avoid sources of bias in key components of the peer review (reviewer participation, selection, engagement, and reporting). A review announcement, released in January

2025, included the general topic and expertise needs (eg, developmental neurotoxicity) but did not identify the specific manuscript. Within 24 hours of the announcement, 152 scientists applied and 5 reviewers were selected automatically by the web platform based on expertise metrics (years of experience, publications, counts of key words relevant to the manuscript). The review panel consisted of qualified scientists (mean [SD] years of experience, 37 [14] years; range, 21-52 years; mean [SD] publication count, 416 [124]; range, 216-550) across 4 countries. Reviewers were automatically verified via a community-driven reputation system based on past interactions with the web platform. To improve robustness of the process, the review was structured as a 3-round modified Delphi format. A $250 honorarium was offered to reviewers to complete their work under an accelerated schedule. Reviewers were masked to each other. In round 1, reviewers worked independently to review the manuscript and answer structured questions that included confidence ratings (0-10 scale, with higher number indicating higher confidence) for components of the manuscript with explanations and confidence in relying on the reported study for making policy decisions by health agencies. In round 2, reviewers worked deliberatively in reviewing and commenting on each other's answers (13 comments submitted). In round 3, the reviewers returned to their round 1 responses, revised as needed after deliberation, and submitted final responses.

Results All 3 rounds of the review were completed within 1 week of panel selection, and honoraria were initiated within 24 hours of completion. A comparison of the panel's independent input from round 1 and their deliberative input in round 3 yielded a decrease in the confidence score of the manuscript (mean [SD] confidence score of 4.2 [2.5] after round 1 was 3.8 [2.2] after round 3) and an improvement in consensus across reviewers (based on Tastle and Wierman[1]), from 0.71 after round 1 to 0.77 after round 3) (**Figure 25-1065**). Confidence in using the study to support decision-making by health agencies was considered low by the review panel (mean [SD] score, 3.0 [1.2]).

Conclusions This pilot study found that rapid (ie, within 1 week) peer reviews can be performed with a process that includes multiple rounds of engagement and collaborative sharing of reviewer ratings and comments in a scalable manner via automation of multiple steps in the process, including ranking of candidate reviewers, contracting (agreement defining scope of work and compensation), scheduling of rounds, messaging, and payments.

Reference

1. Tastle WJ, Wierman MJ. Consensus and dissention: a measure of ordinal dispersion. *Int J Approx Reason*. 2007;45(3):531-545. doi:10.1016/j.ijar.2006.06.024

[1]SciPinion, Bozeman, MT, ckirman@scipinion.com.

Conflict of Interest Disclosures None reported.

Funding/Support This work was self-funded by SciPinion.

Predatory Journals

In-person

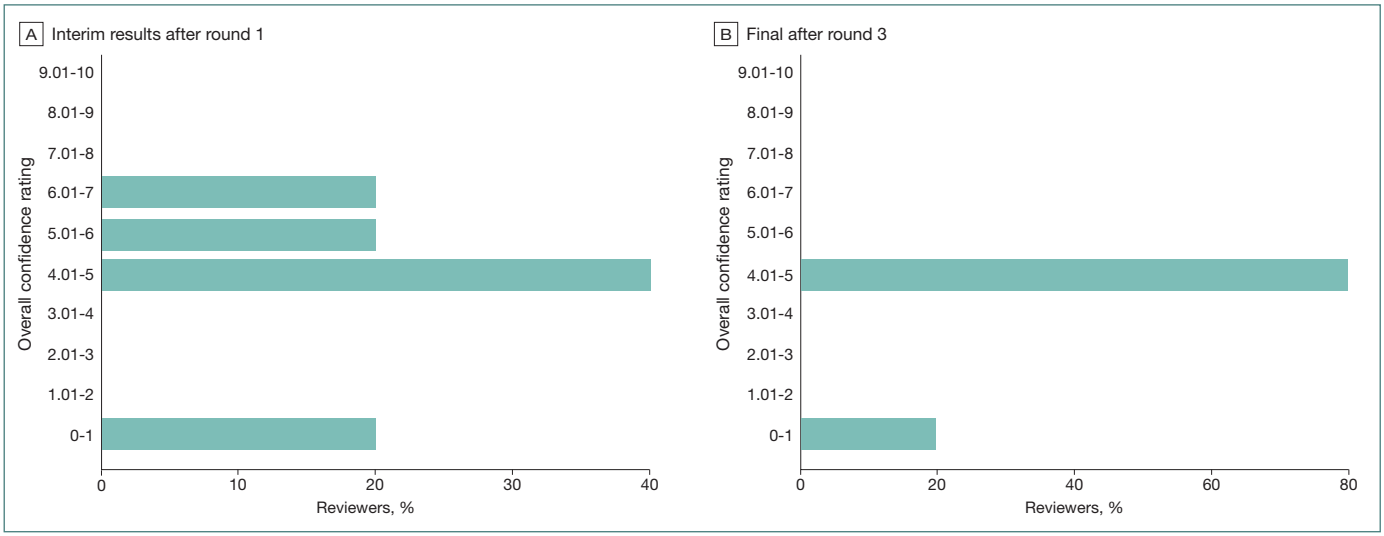
Persistence and Indexing of Predatory Journals and Publishers: A Follow-Up Evaluation of Beall's List

Pravin Bolshete,[1] Madhulika Bolshete,[1] Priyanka Mate[1]

Objective To assess the current status of journals and publishers formerly listed in Beall's List, including their operational status, recent publication activity, and indexing in recognized databases.

Design A systematic evaluation of journals and publishers listed in Beall's List was conducted. The list was accessed from https://beallslist.net/, considering Beall's original list has been defunct since January 2017. The current version of the list is hosted anonymously and was used solely as a tool to access the archived entities listed by Beall, excluding updates made after 2017. Each website link was checked for functionality, and the operational status of the respective journal or publisher was recorded from November to

Figure 25-1065. Reviewer Confidence Scores, 0-10 Scale



December 2024. Among journals that remained active, we further assessed whether they had published articles in 2024 and their indexing status, listed on their website, in databases such as PubMed/Medline, Scopus, Embase, Directory of Open Access Journals, and Web of Science.

Results Of the total number of journals (N = 1310) and publishers (N = 1163) assessed, 617 (47.1%) and 624 (53.7%), respectively, were found to be defunct, with their websites no longer accessible. Among the journals that remained active (n = 693 [52.9%]), 461 (66.5%) had published articles in 2024. Indexing analysis revealed that 154 (22.2%) of these journals mentioned that they are listed in at least 1 recognized academic database.

Conclusions A significant proportion of predatory journals and publishers listed in Beall's List have ceased operations, yet a substantial number remain active and continue to publish research. Some of these journals have managed to be indexed in reputed databases, raising concerns about the persistence and evolving strategies of predatory publishing. The operational status of a journal and its indexing do not confirm or refute its current quality. Some journals may have reformed and implemented peer review and ethical publishing practices. However, we cannot confirm whether active journals have improved, remained predatory, or now operate under misleadingly enhanced credibility. Additionally, although the use of Beall's archived list offers a consistent reference point, the anonymous nature of its current hosting presents limitations. Overall, these findings underscore the need for ongoing scrutiny and awareness required to safeguard academic integrity and prevent researchers from inadvertently submitting their work to predatory journals.

[1]Sqarona Medical Communications, Pune, India, pravinbolshete@gmail.com.

Conflict of Interest Disclosures All authors are affiliated with Sqarona Medical Communications, which provides medical writing services.

Virtual

Inclusion of Randomized Controlled Trials Published by Potentially Predatory Journals in Anesthesiology Systematic Reviews: A Cross-Sectional Study

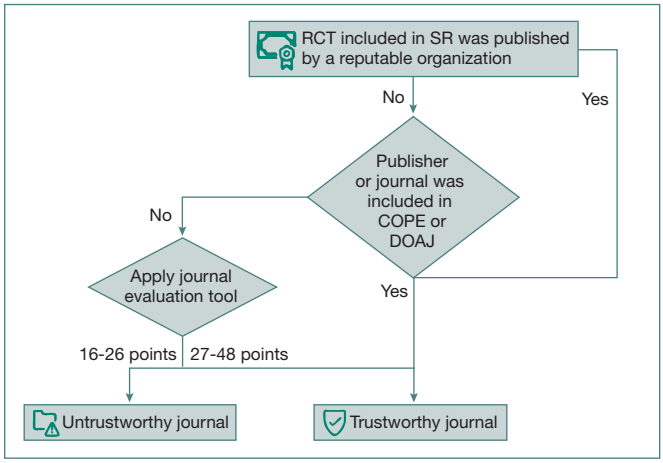
Julián C. Velásquez Paz,[1] Andrés Zorrilla Vaca,[2] Markus Klimek,[3] Jose A. Calvache[1,3]

Objective In recent years, systematic reviews (SRs) and meta-analyses (MAs)—traditionally regarded as reliable sources of causal evidence due to their methodological rigor—have increasingly been compromised by the inclusion of predatory literature, which is marked by misleading information and questionable editorial practices.[1] Despite ongoing efforts,[2] predatory literature continues to pose a major challenge in evidence synthesis. We aimed to assess the degree of inclusion of randomized controlled trials (RCTs) published in potentially predatory journals into SRs with MAs published in leading anesthesiology journals.

Methods We conducted a cross-sectional study in which we included SRs with MAs published between 2021 and 2022 in the top 10 high-ranked anesthesia journals based on the Web of Science's Journal Citation Report for 2021. Exclusion criteria included SRs without MAs or network MAs as well as those including observational studies and nonhuman studies. To determine whether an RCT was published in a potentially predatory journal, 2 authors independently followed a stepwise process: (1) an initial assessment of the publisher's reputation, confirming if it was widely recognized and reputable in the academic publishing community; (2) verification of the journal's indexing in the Committee on Publication Ethics or Directory of Open Access Journals; and (3) application of the Loyola Marymount University Journal Evaluation Tool (JET).[3] Discrepancies between the 2 authors were resolved by consensus. The JET evaluates journals based on key criteria such as transparency of editorial processes, rigor of peer review practices, adherence to ethical standards, presence in reputable indexing databases, archiving practices, and the quality of the journal's website; each aspect is scored using defined indicators based on their credibility and risk of predatory practices (**Figure 25-1165**). We summarized our results using descriptive statistics.

Results A total of 79 SRs with MAs including 1110 RCTs were analyzed, with a median of 10 RCTs per SR. Most studies originated from the US, China, and India (432 [38.9%]). After applying the stepwise approach, 29 (2.6%) of the analyzed RCTs were identified as published in journals classified by our process as untrustworthy; in addition, RCTs published in potentially predatory journals were included in 14 (17.7%) of the 79 studied SRs. Finally, after reviewing Beall's list, www.predatoryjournals.org, and journals excluded from the Web of Science or Scopus since 2017, we found that only 5 (17.2%) of the RCTs published in journals classified as untrustworthy appeared on any of these lists.

Figure 25-1165. Process for Assessing Randomized Clinical Trials Published by Potentially Predatory Journals



COPE indicates Committee on Publication Ethics; DOAJ, Directory of Open Access Journals; RCT, randomized clinical trial; SR, systematic review.

Conclusions This study highlights that RCTs published in potentially predatory journals are being cited in SRs with MAs published in leading anesthesiology journals. Our findings indicate that while lists of potentially predatory journals can be helpful, they are probably not sufficient to fully identify actual predatory journals. Further analysis is needed to determine the impact of RCTs published by potentially predatory journals on the main findings of these MAs.

References

1. Grudniewicz A, Moher D, Cobey KD, et al. Predatory journals: no definition, no defence. *Nature*. 2019;576(7786):210-212. doi:10.1038/d41586-019-03759-y

2. Cukier S, Helal L, Rice DB, et al. Checklists to detect potential predatory biomedical journals: a systematic review. *BMC Med*. 2020;18(1):104. doi:10.1186/s12916-020-01566-1

3. Rele S, Kennedy M, Blas N. Journal Evaluation Tool. Loyola Marymount University. 2017. Accessed July 10, 2024. https://digitalcommons.lmu.edu/librarian_pubs/40

[1]Department of Anaesthesiology, Universidad del Cauca, Popayan, Colombia, jacalvache@unicauca.edu.co; [2]Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA, US; [3]Department of Anesthesiology, Erasmus University Medical Center Rotterdam, the Netherlands.

Conflict of Interest Disclosures None reported.

Preprints

In-person

Preprint Policies in Ecology and Evolutionary Biology Journals

Marija Purgar,[1,2] Edward R. Ivimey-Cook,[3] Antica Culina,[1,4] Joshua D. Wallach[2,5]

Objective Preprints—preliminary research reports that have not yet undergone peer review—are becoming increasingly common across scientific fields.[1] However, little is known about preprint policies in journals that publish ecology and evolutionary biology research.[2]

Design In this cross-sectional analysis, we identified all journals included under the Web of Science 2023 Journal Citation Reports categories of Ecology and Evolutionary Biology. We reviewed journal and publisher websites to explore specific preprint policies, including guidelines regarding the possible locations for posting preprints (eg, on a preprint server), disclosure of preprints at time of manuscript submission and the required location (eg, in the cover letter), timing of preprint posting relative to manuscript submission (eg, before submission to a journal), and any additional actions on publication of the final article (eg, linking the preprint to the published version). The journal-level and publisher-level policies were classified as preprints allowed without restrictions, preprints considered on a case-by-case determination (manuscripts with preprints evaluated by the journal on an individual basis), or preprints prohibited. Journals with no identified policy were categorized as having no preprint policy. We used a Wilcoxon rank sum test to compare 2023 Journal Impact Factors between journals with a preprint policy at either journal- or publisher-level and those without any preprint policy.

Results We identified 230 eligible ecology and evolutionary biology journals published by 69 different publishers, of which 119 (51.7%) included preprint policies in their author guidelines—either through journal-specific policies (n = 109) or by directly referencing their publisher's preprint policies (n = 10). There were 73 additional journals (31.7%) without their own preprint policies, which did not directly reference their publisher's preprint policies but were associated with publishers that had favorable policies. Overall, there were 38 journals (16.5%) without any journal-level or publisher-level preprint policies. Of the 192 journals with either journal-level or publisher-level preprint policies, 191 (99.5%) explicitly allowed preprints and 1 (0.5%) considered preprints on a case-by-case basis. The median (IQR) Journal Impact Factor was higher among journals with a journal-level or publisher-level preprint policy (2.4 [1.7-3.7]) compared with those without any preprint policy (0.6 [0.5-0.9]) ($P < .001$). Among the 109 journals with journal-level preprint policies (**Table 25-1002**), 62 (56.9%) provided information about specific platforms (eg, bioRxiv, SSRN, arXiv), 24 (22.0%) required disclosure of preprints during manuscript submission, and 35 (32.1%) provided specific guidance on the timing of preprint posting relative to manuscript submission.

Conclusions In this study, only one-half of identified ecology and evolutionary biology journals provided preprint policies in their journal-level author guidelines or offered clear instructions regarding the use of preprints. This lack of clarity may lead to confusion among authors about whether posting a preprint before submitting to a peer-reviewed journal conflicts with the journal's policies on prior publication.

References

1. Abdill RJ, Blekhman R. Meta-research: tracking the popularity and outcomes of all bioRxiv preprints. *eLife*. 2019;8:e45133. doi:10.7554/eLife.45133

2. Noble DW, Xirocostas ZA, Wu NC, et al. The promise of community-driven preprints in ecology and evolution. *EcoEvoRxiv*. Preprint posted online June 13, 2024. doi:10.32942/X2SS46

[1]Division for Marine and Environmental Research, Ruđer Bošković Institute, Zagreb, Croatia; [2]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, US, joshua.wallach@emory.edu; [3]School of Biological Sciences, One Health & Veterinary Medicine, University of Glasgow, Scotland, UK; [4]Netherlands Institute of Ecology, Royal Netherlands Academy of Arts and Sciences, Wageningen, the Netherlands; [5]Collaboration for Regulatory Rigor, Integrity, and Transparency, Yale School of Medicine, New Haven, CT, US.

Conflict of Interest Disclosures Marija Purgar, Edward R. Ivimey-Cook, and Antica Culina are members of the *Society for Open, Reliable, and Transparent Ecology and Evolutionary*

Table 25-1002. Summary of Preprint Policy Among 230 Ecology and Evolutionary Biology Journals, With Detailed Characteristics for 109 Journals With Journal-Level Preprint Policies

Category	Journals, No. (%)
All journals (N = 230)	
Journal-level policy	109 (47.4)
Publisher-level policy	83 (36.1)
No policy	38 (16.5)
Journal-level preprint policy (n = 109)	
Explicitly allowed	106 (97.2)
Case-by-case determination	3 (2.8)
Preprint policy location[a]	
Journal author guidelines	89 (81.7)
Journal policies	15 (13.8)
Other	5 (4.6)
Acceptable preprint posting locations[b]	
Preprint servers	79 (72.5)
Author's homepage	12 (11.0)
Institutional repositories	8 (7.3)
Other	23 (21.1)
Not specified	11 (10.1)
Reference to specific platforms[b]	
Yes	62 (56.9)
bioRxiv	25 (40.3)
SSRN	23 (37.1)
arXiv	14 (22.6)
Other	37 (59.7)
No	47 (43.1)
Guidance on disclosure of preprints at time of manuscript submission	
Yes	24 (22.0)
Cover letter	11 (45.8)
Within manuscript	2 (8.3)
Within submission portal	2 (8.3)
Not specified	9 (37.5)
No	85 (78.0)
Guidance on timing of preprint posting relative to manuscript submission	
Yes	35 (32.1)
Any time	20 (57.1)
Prior to submission	11 (31.4)
Any time prior to acceptance	4 (11.4)
No	74 (69.7)
Guidance on linking preprint to published article	
Yes	48 (44.0)
No	61 (56.0)

[a]Preprint policy location was categorized as follows: journal author guidelines includes any reference to a journal's preprint policies within its author guidelines; journal policies refer to broader editorial, publishing, or ethics policy pages where the preprint policy was mentioned; and other covers other locations, such as submission checklists and terms and conditions.

[b]Percentages exceed 100% because some journals allow preprints to be posted in multiple locations (such as preprint servers, authors' homepages, and institutional repositories) and on multiple servers (such as bioRixv and arXiv). Each location or server is counted separately in the analysis, meaning a single journal may be represented across multiple categories.

Biology; Marija Purgar and Edward R. Ivimey-Cook are on the current Board of Directors. Joshua D. Wallach reports funding from Arnold Ventures to the Yale Collaboration for Regulatory Rigor, Integrity, and Transparency, the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award 1K01AA028258, Johnson & Johnson (through the Yale Open Data Access Project), and the US Food and Drug Administration. No other disclosures were reported.

Funding/Support Marija Purgar was funded by the Croatian Science Foundation project number DOK-2021-02-6688. Marija Purgar gratefully acknowledges the financial support for this research by the Fulbright US Student Program, which is sponsored by the US Department of State and the Croatian-American Fulbright Commission. Antica Culina was funded by the Croatian Science Foundation project number IP-2022-10-2872.

Additional Information Marija Purgar is a co–corresponding author (mpurgar@irb.hr).

Authors' Journeys From Preprints With *The Lancet* on SSRN to Publication

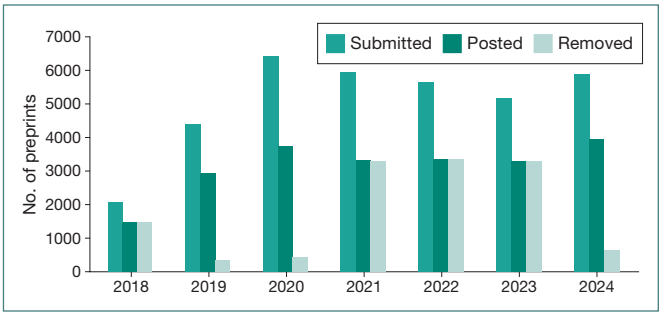
Sherrie L. Kelly,[1] Clare F. Stone,[2] Catherine Fiscus,[2] Ashlie Jackman-Juler,[2] Miriam Lewis Sabin[1]

Objective Previous research has highlighted the importance of preprints in increasing the transparency and rapid dissemination of scientific communication. This study examines publication rates of manuscripts submitted to Preprints With *The Lancet* (PPwTL), a collaboration between The Lancet Group and SSRN, Elsevier's preprint server.

Design This quality improvement study had 3 phases. First, we assessed the number of preprints submitted, posted, and removed from PPwTL from January 1, 2018, to December 1, 2024. Second, we focus on the number of preprints submitted in this period where a version of record (VOR) can be found. Finally, we surveyed corresponding authors of preprints without publications at Elsevier to explore potential barriers to publication.

Results From 2018 to 2024, among The Lancet Group's 24 journals, 35,703 preprints were submitted to PPwTL with annual submissions increasing by 183% from 2089 in 2018 to 5915 in 2024 (**Figure 25-1109**). The annual acceptance rate declined slightly, dropping to 67% in 2024 (3987 accepted of 5915 submitted) from 72% in 2018 (1509 of 2089), while the rate of removals at the authors' request reached 17% in 2024 (686 removed of 3987 posted), compared with 9% in 2018 (130 of 1509). Throughout this period, the average time from submission to posting was 2 days, including our screening processes. We ran an internal report to identify how many of the 35,703 preprints that were submitted have a VOR in

Figure 25-1109. Preprints Submitted to the Collaboration Preprints With *The Lancet*



Science Direct (Elsevier publications). We identified 18,543 (51.9%) submissions with an Elsevier VOR and 17,160 (48.0%) with no VOR in Science Direct. To identify publications from PPwTL submissions that occur in non-Elsevier journals, we cross-checked the 17,160 papers with an internal Fate of Manuscript FORMS report, which lists VORs outside of Elsevier. This report revealed 2639 of the initial 35,703 submissions (7.4%) had non-Elsevier publications in the period 2020 to 2024. The combined numbers of papers with a VOR (18,543 in Elsevier journals and 2,639 outside of Elsevier) is 21,182 and represents an overall publication rate of 56.3% for PPwTL submissions. This number is likely to be higher, because the FORMS report covers publications from 2020 to 2024 and the initial submissions data cover 2018 to 2024. Of the 2639 publications in non-Elsevier journals, 507 (19.2%) were published in Springer Nature journals; 145 (5.5%) in PLOS journals; 146 (5.5%) in BMJ journals, and 25 (0.9%) in JAMA Network journals. Findings from the ongoing author survey are anticipated in the coming months.

Conclusions It is important to understand the impact of preprints on the scholarly record, and it's encouraging to note the increased acceptance of preprints. A significant number of PPwTL manuscripts that do not get published in Elsevier journals nevertheless are published elsewhere, benefitting the scholarly community and strengthening the preprint/journal landscape.

[1]The Lancet Group, Elsevier, London, UK; [2]SSRN, Elsevier, Rochester, NY, US, c.stone@elsevier.com.

Conflicts of Interest Disclosures None reported.

Acknowledgments The authors express their gratitude to current and former members of the Preprints With*The Lancet*-SSRN Working Group, to Aschwin Wijnsma (Elsevier) who assisted with the FORM report used to examine the fate of manuscripts in this study, and to authors who contributed to Preprints With *The Lancet* and responded to the survey with consent.

Preregistration of Studies

In-person

Impact of ICMJE Trial Registration Policy at 20 Years

Julianne T. Nelson,[1] Tony Tse,[1] Swapna Mohan,[1] Yvonne Puplampu-Dove[1]

Objective The 2004 International Committee of Medical Journal Editors (ICMJE) policy required prospective registration of clinical trials as a condition for publication. Prior work reported widespread inconsistencies at earlier time points, including incomplete registration,[1] retrospective or unregistered trials,[2] and discrepancies between registered and published primary outcome measures.[3] This analysis aims to build on previous work, now 20 years post-ICMJE policy, and expands to include the global landscape of trial publication and registration.

Design This cross-sectional analysis assessed a convenience sample of primary trial results publications first indexed in PubMed from July to September 2024 that was randomly selected by funding category to reflect a distribution of 40% industry and 60% nonindustry.[2] Each article was manually reviewed by 1 author. The journal title, first publication date, and any unique identifier from an ICMJE-recognized registry listed in the abstract or body text were extracted. Publications without trial registry information had their status confirmed via advanced search within the WHO International Clinical Trials Registry Platform (ICTRP). From a random subset of 100 sampled publications listing registry identifiers, again selected randomly for a 40% to 60% funding distribution, 1 author extracted first registration and study start dates from corresponding registration records to determine timing of registration (prospective or retrospective). The author also compared primary outcome measures (POMs) listed in publications and records. The first outcome measure mentioned was used for publications without explicitly identified POMs. Discrepancy codes were adapted from prior work.[3] Registered POMs with vague terms (eg, safety) were coded as "unclear." All data were reviewed independently by a second author and differences resolved by discussion.

Results Of 434 overall sampled articles from 322 journal titles on October 17, 2024, 82.7% (359 of 434) disclosed registration identifiers, of which 212 (59.1%) were listed in abstracts. Of 17 represented registries, ClinicalTrials.gov (69.6% [250 of 359]), ChiCTR (7.0% [25 of 359]), and ANZCTR (3.6% [13 of 359]) were cited most frequently. Among the subset of 100 trials, 23% were retrospectively registered (**Table 25-0939**). Among the extracted 135 POMs, 2.2% (3 of 135) were coded as unclear and 18.5% (25 of 135) as discordant: 52.0% (13 of 25) due to published POMs missing from registration records and 28.0% (7 of 25) due to published POMs registered as secondary outcome measures. No substantive differences between trials funded by industry and nonindustry sources were observed.

Conclusions Twenty years after the ICMJE policy, nearly one-fifth of sampled publications did not disclose registration identifiers. Of those that did, 40% of identifiers were not displayed in abstracts (inaccessible in PubMed). Among the subset, almost one-quarter were registered retrospectively. Of published POMs, approximately one-fifth were discordant with registered information. These findings do not differ widely from previous work of 10 years ago and suggest that the policy goals of full transparency and accountability through trial registration have yet to be reached.

References

1. Viergever RF, Karam G, Reis A, Ghersi D. The quality of registration of clinical trials: still a problem. *PLoS One*. 2014;9(1):e84727. doi:10.1371/journal.pone.0084727

2. Gopal AD, Wallach JD, Aminawung JA, et al. Adherence to the International Committee of Medical Journal Editors' (ICMJE) prospective registration policy and implications for outcome integrity: a cross-sectional analysis of trials published in high-impact specialty society journals. *Trials*. 2018;19(1):448. doi:10.1186/s13063-018-2825-y

Table 25-0939. Characterization of Subset of Primary Trial Results Publications by Funding Source (N = 100)

Attribute	By funding source,[a] No. (%)		
Timing of registration	Industry (n = 40 publications)	Nonindustry (n = 60 publications)	Total (N = 100 publications)
Prior to or on start date[b]	33 (82.5)	44 (73.3)	77 (77.0)
After start date	7 (17.5)	16 (26.7)	23 (23.0)
Within 3 mo of start date	4 (57.1)	6 (37.5)	10 (43.5)
More than 3 mo after start date	3 (42.9)	10 (62.5)	13 (56.5)
Comparison: POMs	Industry (n = 53 POMs)	Nonindustry (n = 82 POMs)	Total (N = 135 POMs)
Concordant	41 (77.4)	66 (80.5)	107 (79.3)
Discordant	12 (22.7)	13 (15.9)	25 (18.5)
Not registered	8 (66.7)	5 (38.5)	13 (52.0)
Registered as secondary outcome measure	1 (8.3)	6 (46.2)	7 (28.0)
Registered as POM but with different time frame	3 (25)	2 (15.4)	5 (20.0)
Unclear—registered POM vague	0	3 (3.7)	3 (2.2)

Abbreviation: POM, published and registered primary outcome measure.

[a]Prior work[2] reported samples of primary trial results publications with a distribution of approximately 40% industry funding and observed differences in trial registration attributes by funding source. This analysis used a selective sampling approach to achieve a similar distribution in a convenience sample (ie, 40% listing at least 1 industry funding source but no concurrent funding from the National Institutes of Health or other governmental agencies) to explore such differences in an updated sample and using different measures.

[b]If start date was not available in the trial registry, the date provided as study start in the publication (if available) was used.

3. Turner EH, Mulder RT, Rucklidge JJ. Is mandatory prospective trial registration working? An update on the adherence to the International Committee of Medical Journal Editors guidelines across five psychiatry journals: 2015-2020. *Acta Psychiatr Scand*. 2021;144(5):510-517. doi:10.1111/acps.13353

[1]The National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, US, julianne.nelson@nih.gov.

Conflict of Interest Disclosures None reported.

Funding/Support This work 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 funder 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.

Disclaimer The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the National Institutes of Health.

Additional Information Swapna Mohan and Yvonne Puplampu-Dove reported that they conducted this work under contract with ICF International Inc.

Effectiveness of Preregistration in Psychology

Olmo R. van den Akker,[1,2] Marcel A. L. M. van Assen,[1] Marjan Bakker,[1] Jelte M. Wicherts[1]

Objective Preregistration,[1] the practice in which researchers publish their hypotheses, study design, and/or analysis plan before collecting or analyzing their data, has become more and more established in the scientific ecosystem even though the evidence that it prevents bias has been lacking. This set of studies was conducted to assess the effectiveness of preregistration to reduce bias in psychology.

Design In the first study, conducted from 2021 to 2024, we assessed the consistency of preregistrations and their corresponding publications (published in 40 psychology journals). These preregistered publications were published and awarded a preregistration badge by a journal or a preregistration challenge prize by the Center for Open Science from 2017 to 2020. We also assessed whether any inconsistencies were disclosed by the authors. In the second study, we compared a subset of preregistration-publication pairs for which a preregistered statistical result could be extracted with a set of nonpreregistered studies (from 37 journals) that were selected for comparability via the Web of Science related records function. Using multilevel regressions, we examined whether preregistration was associated with *P* hacking, operationalized as a lower proportion of statistical results and smaller effect sizes. We also looked at associations with other research characteristics, such as the number of citations, sample sizes, and the frequency of statistical errors (assessed via Statcheck[2]).

Results The sample included 300 preregistration-publication pairs as well as a subset of 193 preregistration-publications pairs matched with 193 nonpreregistered studies. We found inconsistencies for 6 important study elements (operationalizations of measured, manipulated, and dependent variables, data collection procedures, statistical model, and inference criteria [**Table 25-0947**]). The comparison between preregistered and nonpreregistered studies did not indicate an association between preregistration and statistical significance (β = 0.01; 99% CI, −0.56 to 0.59; *P* = .96), effect sizes (β = −0.04; 99% CI, −0.12 to 0.04; *P* = .18), and statistical errors (β = −1.19; 95% CI, −2.51 to 0.13; *P* = .08). However, preregistered studies typically had higher sample sizes (959.6 vs 536.6;

Table 25-0947. Preregistration-Publication Consistency for the 6 Major Study Elements

Major study elements[a]	Consistency, No. (%)		
	Yes	No	NA[b]
Measured variable (N = 164)	91 (55)	28 (17)	45 (27)
Manipulated variable (N = 218)	146 (67)	64 (29)	8 (4)
Dependent variable (N = 218)	131 (60)	31 (14)	56 (26)
Data collection procedure (N = 300)	84 (28)	40 (13)	176 (59)
Statistical model (N = 300)	135 (45)	44 (15)	121 (40)
Inference criteria (N = 300)	94 (31)	196 (65)	10 (3)

Abbreviation: NA, not applicable.

[a]The total numbers for the measured variable, manipulated variable, and dependent variable are not 300 because not every study contained all these variables.

[b]NA refers to situations where it proved impossible to compare the study elements because insufficient information was provided in either the preregistration, the publication, or both.

β = 0.45; 99% CI, 0.14 to 0.76; *P* < .001) and more citations (18.3 vs 15.1; β = 0.20; 95% CI, 0.01 to 0.40; *P* = .04).

Conclusions Ideally, publications would be concordant with their preregistrations and any deviations would be reported and explained. In these studies, this ideal was often not met. Preregistration was associated with higher statistical power and more subsequent citations, but it did not appear to prevent *P* hacking. The studies were limited by their focus on psychology, and generalizations to other fields should therefore be avoided (although economics and political science had comparable results[3]).

References

1. Rice DB, Moher D. Curtailing the use of preregistration: a misused term. *Perspect Psychol Sci*. 2019;14(6):1105-1108. doi:10.1177/1745691619858427

2. Nuijten MB, Hartgerink CH, Van Assen MA, Epskamp S, Wicherts JM. The prevalence of statistical reporting errors in psychology (1985-2013). *Behav Res Methods*. 2016;48:1205-1226. doi:10.3758/s13428-015-0664-2

3. Ofosu GK, Posner DN. Pre-analysis plans: an early stocktaking. *Perspect Polit*. 2023;21(1):174-190. doi:10.1017/S1537592721000931

[1]Tilburg University, Tilburg, The Netherlands, ovdakker@gmail.com; [2]QUEST Center for Responsible Research, Berlin Institute of Health at Charité, Berlin, Germany.

Conflict of Interest Disclosures None reported.

Funding/Support The work in this presentation was supported by a Consolidator Grant (IMPROVE) from the European Research Council (grant number 726361).

Additional Information Both studies presented in this abstract were preregistered: the assessment of preregistration effectiveness at https://osf.io/qbhsv, and the comparison of preregistered with nonpreregistered studies at https://osf.io/7w5g2.

Registration of Observational Studies of Interventions: Prevalence, Characteristics, and Journal Policies

Cecilie Jespersen,[1,2] Zexing Song,[3] An-Wen Chan,[3,4] Asbjørn Hróbjartsson[1,2]

Objective Observational studies of interventions use causal inference to assess the impact of interventions on health-related outcomes.[1] Despite concerns about reporting bias, observational studies are not subject to the same registration requirements as clinical trials.[2,3] We aimed to determine the prevalence of registration among published observational studies of interventions, assess the association between registration and study characteristics, analyze journal registration policies, and explore authors' and editors' attitudes about registration.

Design We conducted a meta-epidemiologic cross-sectional study triangulating data from 4 sources. First, we searched PubMed for observational studies published in 2023. Eligible studies were cohort or case-control studies with a control group that assessed causal effects of health interventions. Corresponding registration information was collected. Second, authors of included studies were surveyed to explore reasons for and barriers to registration. Third, editorial policies were sampled from 40 journals: 20 sample-representative journals and the journals ranked in the top 20 in Journal Citation Reports by 2023 Journal Impact Factor across 8 specialty categories. Fourth, 1 editor per journal was invited to share their perspectives on registration. Primary outcomes were the prevalence of registered observational studies of interventions published in 2023 and the estimated association between registration and study characteristics, assessed by multivariable logistic regression. Sample size was estimated based on an expected 15% registration rate.

Results Among 1100 screened studies, 200 were included: 69 and 128 cohort studies with prospective and retrospective data collection, respectively, and 3 case-control studies. In total, 28 (14%) were registered, and 17 of these (61%) were prospectively registered (<1 month of their start date) (**Table 25-1069**). Prospective design and protocol availability were positively associated with registration (retrospective vs prospective cohort: odds ratio [OR], 0.19 [95% CI, 0.07-0.54]; *P* = .002; no public protocol vs public protocol: OR, 0.04 [95% CI, 0.01-0.23]; *P* < .001). The survey response rate was 23% (46 responses); 60% of authors supported registration, although many only when registration was deemed relevant. Identified barriers included lack of journal requirements for registration (56%) and limited resources (62%). None of the journal policies explicitly required registration of observational studies of interventions, while 12 (30%) encouraged it. Journals that encouraged registration had a higher 2023 Journal Impact Factor and more frequently encouraged public protocols. Editors had divergent opinions on registration. While some considered it to be worthwhile, just as many questioned the added value.

Conclusions Only 1 in 7 contemporary observational studies of interventions were registered, although more often in cohort studies with prospective data collection and studies with a publicly available protocol. Authors identified the lack of journal requirements to registration as a key registration barrier, and only one-third of journals had supportive policies. Clearer guidance and journal policies on registration relevance (discriminating hypothesis-testing and hypothesis-generating studies) may reduce the risk of reporting biases in observational studies of interventions.

References

1. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA*. 2022;328(24):2446-2447. doi:10.1001/jama.2022.21383

2. Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? *CMAJ*. 2010;182(15):1638-1642. doi:10.1503/cmaj.09225

3. Leducq S, Zaki F, Hollestein LM, et al. The majority of observational studies in leading peer-reviewed medicine

Table 25-1069. Main Characteristics of 200 Observational Studies of Interventions by Registration Status

Characteristic	Studies[a]		
	Prospectively registered (n = 17)	Retrospectively registered (n = 11)	Not registered (n = 172)
Design			
Prospective cohort	12 (71)	9 (82)	48 (28)
Retrospective cohort	5 (29)	2 (18)	121 (70)
Case-control	0	0	3 (2)
Study site			
Single center	4 (24)	6 (55)	100 (58)
Multicenter	13 (76)	5 (45)	72 (42)
Funding			
Industry or mixed	5 (30)	2 (18)	19 (11)
Nonindustry	6 (35)	7 (64)	51 (30)
None	2 (12)	2 (18)	68 (39)
Not reported	4 (20)	0	34 (20)
Sample size, median (IQR)	338 (138-451)	538 (208-4729)	286 (92-2005)
Country of PI			
Europe	9 (53)	5 (45)	72 (42)
Asia	4 (25)	2 (18)	50 (29)
North America	4 (24)	3 (27)	38 (22)
Other	0	1 (9)	12 (7)
Primary outcome			
Statistically significant	10 (59)	8 (73)	130 (76)
Not statistically significant	7 (41)	3 (27)	42 (24)
Protocol			
Publicly available	6 (35)	5 (45)	3 (2)
None provided	11 (65)	6 (55)	169 (98)
Ethical approval			
Approved	17 (100)	11 (100)	152 (88)
Not approved[b]	0	0	20 (12)
2023 JIF, median (IQR)	4.9 (2.5-8.7)	4.9 (2.9-6.5)	2.6 (1.9-3.8)

Abbreviations: JIF, Journal Impact Factor; PI, principal investigator.
[a]Data are presented as number (percentage) of studies unless otherwise specified.
[b]Explicitly not approved or no approval mentioned in the article.

journals are not registered and do not have a publicly accessible protocol: a scoping review. *J Clin Epidemiol*. 2024;170:111341. doi:10.1016/j.jclinepi.2024.111341

[1]Cochrane Denmark & Centre for Evidence-Based Medicine Odense (CEBMO), University of Southern Denmark, Odense, Denmark, ceciliejespersen@health.sdu.dk; [2]Open Patient data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark; [3]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; [4]Women's College Research Institute, Dept. of Medicine, University of Toronto, Toronto, Ontario, Canada.

Conflict of Interest Disclosures An-Wen Chan 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.

Acknowledgments We thank all researchers who participated in the author survey for their valuable contribution to the findings of this study.

International Registered Reports Identifiers (IRRIDs): 7 Years of Experiences

Gunther Eysenbach[1]

Objective Registered Reports (RRs) refer to the publication of a study that is published in 2 stages: a protocol (RR stage 1 or RR1) and a results paper (RR2). Some journals have adopted the RR system and guarantee acceptance of subsequent results articles published in the same journal after the protocol is peer reviewed. However, in a distributed open science ecosystem, protocols may already be peer reviewed and published elsewhere, but no standardized system exists to link protocols to subsequent results articles and vice versa. We implemented RRs across the publisher portfolio, with 1 dedicated journal to peer review and publish protocols, and the publisher guaranteeing acceptance of RR2s in 1 of its other journals independently based on whether the results were negative or positive. We propose a machine- and human-readable mechanism to link RR2s with RR1s to assist in peer review and to enhance transparency, accountability, and reproducibility.

Design In 2018, we proposed and implemented a cross-journal, DOI-based, persistent identifier called an International Registered Report Identifier (IRRID) published in article abstracts. An RR2 references an RR1 using an identifier that is based on the DOI of the protocol. For example, RR2-10.2196/24264 indicates that the publication is a results article for a protocol that was previously published under the DOI 10.2196/24264.[1] Protocols that are peer reviewed contain the IRRIDs in the format [DE|P]RR1-[DOI], where [DOI] is the DOI of the protocol itself, and DE or P qualifiers indicate whether the protocol was written before [P] or after [DE] data collection. On submission of a protocol, authors were asked if the protocol was submitted before or after data were collected. Authors were incentivized to register their protocol by being offered a 20% discount on the Article Processing Charge on subsequent results articles.

Results A total of 3995 articles were published with IRRIDs between 2018 and February 13, 2025, of which 3240 (81%) were protocols (RR1). Among the protocols, 2151 (66%) were published when data already existed (DERR1), and 917 (28%) were published before data were collected (PRR1). A total of 732 results articles had an RR2 identifier indicating previous protocol publication, although these protocols were not always peer reviewed (eg, OSF or *BMJ Open*). A third-party audit found that outcome switching and undeclared deviations from the protocol still remain a problem.[2]

Conclusions We propose that other journals adopt IRRIDs to help identify protocol and results article pairs that together form RRs across journals (https://irridregistry.org/).

References

1. What is an International Registered Report Identifier (IRRID)? JMIR Publications Knowledge Base and Help Center. Accessed February 13, 2025. https://support.jmir.org/hc/en-us/articles/360003797672-What-is-an-International-Registered-Report-Identifier-IRRID

2. Anthony N, Tisseaux A, Naudet F. Published registered reports are rare, limited to one journal group, and inadequate for randomized controlled trials in the clinical field. *J Clin Epidemiol*. 2023;160:61-70. doi:10.1016/j.jclinepi.2023.05.016

[1]JMIR Publications, Toronto ON, Canada, geysenba@gmail.com.

Conflict of Interest Disclosure Gunther Eysenbach has an equity stake in and receives a salary from JMIR Publications.

Virtual

Outcome Switching in Observational Studies of Interventions: Comparison of Registration Records and Published Articles

Zexing Song,[1,2] Cecilie Jespersen,[3,4] Asbjørn Hróbjartsson,[3,4] S. Joseph Kim,[1,5,6] Rob Fowler,[1,5] Peter C. Austin,[1,6] An-Wen Chan[1,2,5]

Objective Outcome switching between study design and reporting is a potential source of bias in observational studies, but there is a paucity of evidence as to its frequency. We aimed to estimate the prevalence of outcome switching in observational studies of interventions (defined as controlled cohort studies investigating the causal effects of interventions on health-related outcomes). Secondary aims included assessing the completeness of prespecification of primary outcomes and factors associated with outcome switching.

Design This meta-epidemiological study involved longitudinal analyses of observational studies of interventions prospectively registered on ClinicalTrials.gov within 1 month of their study start date between 2014 and 2016 that had results published in a peer-reviewed journal. We screened registry records from January through December 2024 to create the study sample and completed outcome data extraction and analysis from January through April 2025. Complete outcome prespecification required explicit definition in the registry of the measurement variable, analysis metric, method of aggregation (the statistic to summarize the outcome within each group), and time point of the outcome. We evaluated outcome switching by identifying discrepancies in the primary outcomes between the registry and published articles, including omission (prespecified primary outcomes not reported), downgrading (prespecified primary outcomes reported as nonprimary), upgrading (prespecified nonprimary outcomes reported as primary), and introduction of new primary outcomes not listed in the registry. We considered outcome switching to favor statistically significant results if a new statistically significant primary outcome was introduced or upgraded or a nonsignificant one was downgraded. We performed multivariable logistic regression to estimate the association between study characteristics and outcome switching.

Results We screened 9965 registry records labelled as observational studies and included 127 eligible studies with results published between January 2015 and October 2024. Only 23 studies (18%) completely prespecified their primary outcome in the registry, and the method of aggregation was the least commonly defined element (33 [26%]). Outcome switching was found in 60 studies (47%), and only 1 of these studies (2%) provided a rationale for the changes. The most common discrepancy was omission (32 [25%]) followed by downgrading (30 [24%]), introduction of new primary outcomes (23 studies with 29 new primary outcomes [18%]), and upgrading (2 [2%]). New primary outcomes differed most commonly between the registry and published articles in the measurement variable (21 of 29 [72%]) and time point (15 of 29 [52%]). Among 54 studies that had discrepancies not limited to omitted primary outcomes, statistically significant results were favored in 80% (43 of 54). No study characteristics were significantly associated with outcome switching (**Table 25-0874**).

Conclusions Unreported outcome switching and inadequate outcome prespecification were common in observational studies of interventions. These findings underscore the need for improved registration practices and greater transparency

Table 25-0874. Association Between Study Characteristics and Outcome Switching

Study characteristic	Studies		AOR (95% CI)[c]
	Total (N = 127)[a]	Published with outcome switching[b]	
Exposure type			
Drug or biologic	63	28 (44)	0.69 (0.31-1.50)
Other	64	32 (50)	1 [Reference]
Planned sample size, median (IQR)[d]	300 (100-1000)	330 (100-950)	1.02 (0.84-1.24)
Design			
Retrospective cohort	36	18 (50)	1.44 (0.58-3.63)
Prospective cohort	91	42 (46)	1 [Reference]
Type of funder			
Industry involved	42	21 (50)	1.34 (0.59-3.08)
Nonindustry	85	39 (46)	1 [Reference]
Journal impact factor per 5-point increase, median (IQR)	4 (3-6)	4 (3-7)	0.99 (0.83-1.16)
Statistical significance of primary outcomes[e]			
Significant	101	49 (48)	1.14 (0.43-3.10)
Not significant	23	10 (43)	1 [Reference]

Abbreviation: AOR, adjusted odds ratio.

[a]Data are presented as number of studies unless otherwise indicated.

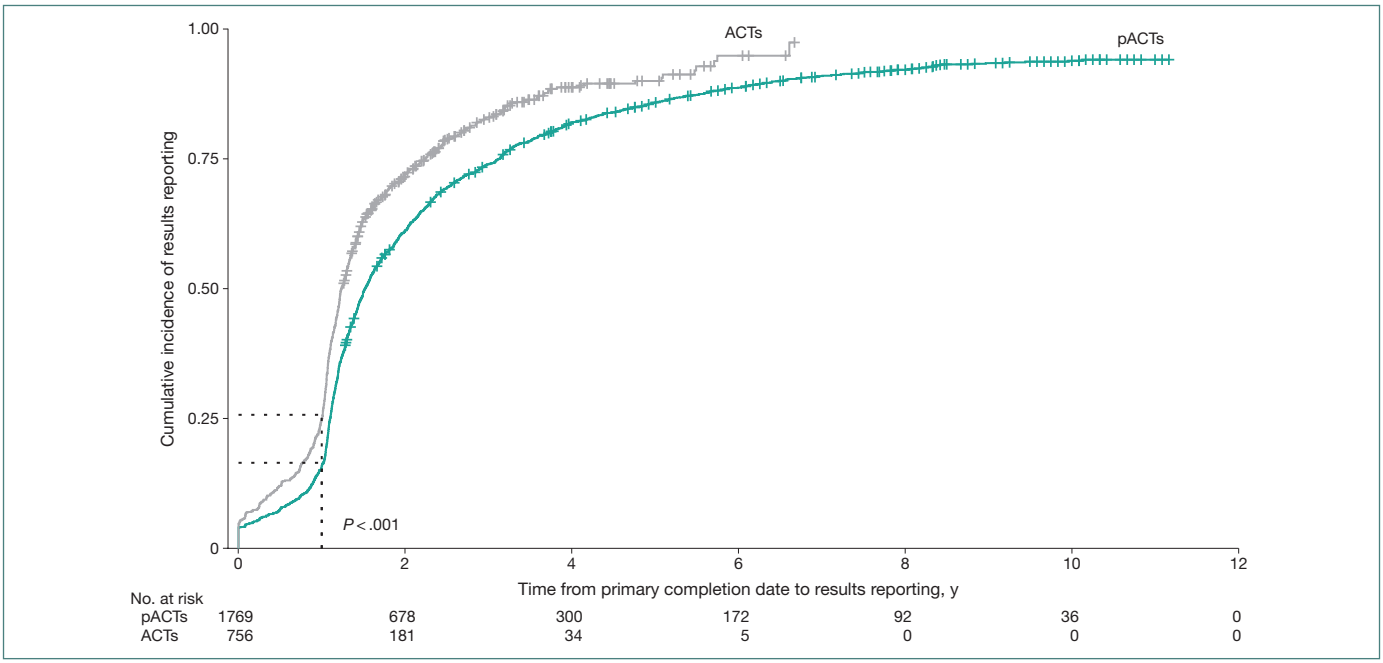
[b]Data are presented as number (row percentage) of studies unless otherwise indicated.

[c]From multivariable logistic regression adjusting for listed study characteristics.

[d]Planned sample size was log transformed (natural log) before inclusion in the regression model to account for the highly skewed and sparse distribution of the data. Odds ratios reflect the change in odds per 1-unit increase in the log-transformed sample size.

[e]Based on reported outcomes in the publication; 3 studies without any statistical tests were excluded.

Figure 24-0810. Time to Results Reporting by Clinical Trial Type



ACT indicates applicable clinical trial; pACT, probable applicable clinical trial.

to better understand the risk of bias in observational research.

[1]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada, zexing.song@mail.utoronto.ca; [2]Division of Dermatology, Women's College Research Institute, Women's College Hospital, Toronto, Ontario, Canada; [3]Cochrane Denmark & Centre for Evidence-Based Medicine Odense (CEBMO), University of Southern Denmark, Odense, Denmark; [4]Open Patient data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark; [5]Department of Medicine, University of Toronto, Toronto, Ontario, Canada; [6]ICES, Toronto, Ontario, Canada.

Conflict of Interest Disclosures An-Wen Chan 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.

Quality of Reporting

In-person

Assessing the Quality and Timeliness of Results Reporting for Clinical Trials on Antimicrobial Agents

Megan Curtin,[1] Allisun Wiltshire,[1] Brix Kowalski,[2] Maximilian J. Siebert[3]

Objective Antimicrobial resistance (AMR) is one of the leading causes of death globally. Timely and complete reporting of clinical trials involving antimicrobial agents (AMAs) is essential to evaluate the safety and efficacy of potential therapies for public use. This study investigates the quality of results reporting of applicable clinical trials (ACTs) and probable ACTs (pACTs) for AMAs. ACTs are interventional studies (excluding phase 1) regulated by the US Food and Drug Administration with at least one site based in the United States. pACTs adhere to the same criteria; however, these studies were initiated prior to January 18, 2017, when the Final Rule came into effect. The rule strengthened reporting obligations by requiring a designated responsible party for results submission to ClinicalTrials.gov and provided clear ACT designation criteria with stricter compliance requirements. We assess how this regulatory change affected the reporting of trials containing AMAs, focusing on timeliness and study characteristics.

Design We extracted data from ClinicalTrials.gov for trials involving AMAs with primary completion dates between May 1, 2013, and May 1, 2023. We analyzed the time from primary completion to results reporting and estimated the hazard ratio to compare timeliness between ACTs and pACTs. Additionally, we assessed delays in reporting across different study types and funding sources.

Results Our search resulted in 2629 trials. We excluded 104 because they only included agents with nonantimicrobial properties. We found 1796 pACTs (71.1%; 95% CI, 69.3%-72.9%) and 756 ACTs (29.9%; 95% CI, 28.2%-31.8%). Among the 2525 trials sampled, 2249 trials (89.1%; 95% CI, 87.8%-90.2%) were reported on ClinicalTrials.gov. There was a median (IQR) of 0.5 years (0.2-1.5) between 1 year after the primary completion date and the results submissions across all studies. The median (IQR) time lag for late reporting for ACTs was 0.3 years (0.1-0.9), while it was 0.6 years (0.2-1.8) for pACTs. Overall, 81.3% (95% CI, 79.7%-82.3%) of trials were reported late (75.0% of ACTs vs 83.6% of pACTs). Our analysis showed that ACTs were more likely to report results earlier than pACTs, with a hazard ratio of 1.4 (95% CI, 1.3-1.5) (**Figure 24-0810**). Regarding funding sources, trials supported by philanthropic foundations or private donors had the slowest reporting (median [IQR] delay, 0.7 [0.3-1.5] years), whereas government-funded trials had the most timely reporting (median [IQR] delay, 0.4 [0.2-1.0] years).

Conclusions Studies including AMAs designated as ACTs demonstrated higher rates of reporting compliance and shorter delays in the reporting of overdue results. While this analysis provides initial insights, limitations related to timeline and sample scope suggest that broader investigations are needed to fully evaluate the policy's impact.

[1]University of California, Berkeley, CA, US; [2]University of Santa Cruz, Santa Cruz, CA, US; [3]Harvard-MIT Center for Regulatory Science, Harvard Medical School, Boston, MA, US, maximiliansiebert91@gmail.com.

Conflict of Interest Disclosures None reported.

Additional Information The protocol for this study was registered at https://osf.io/zp8ug/.

Reporting Study Designs in Korean Medical Journal Articles

Soo Young Kim,[1] Sue Kim,[2] Hyun Jung Yi[3]

Objective This study evaluated whether study designs were reported in papers published in Korean medical journals and whether the reported study designs were appropriate. Additionally, factors influencing such reporting and appropriateness were investigated.

Design A cross-sectional study was conducted in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. A total of 600 articles, representing 5% of the 12,000 articles indexed in KoreaMed (www.koreamed.org, a comprehensive search portal managed by the Korean Association of Medical Journal Editors that includes 288 member journals in the fields of medicine, dentistry, nursing, nutrition, and veterinary medicine) were randomly selected for analysis between January and December 2023. The inclusion criteria specified human participant studies published in either Korean or English. Articles classified as reviews, letters, editorials, or case reports were excluded. These exclusions were applied during the initial random selection process. The primary outcomes assessed were (1) whether the research design was stated in the title, abstract, or methods section; (2) whether the research design was appropriate, as evaluated using the Design Algorithm for Medical Literature on Intervention[1]; and (3) whether the journal submission guidelines referenced reporting guidelines, including mention of the EQUATOR Network.[2] Two independent reviewers evaluated the full texts of the selected articles. Any discrepancies in assessment were resolved through discussion until a consensus was achieved. A χ^2 test was used to examine associations between the reporting study design and (1) the inclusion of reporting guidelines in journal submission policies and (2) references to the EQUATOR Network.

Results Of the 600 articles obtained through a 5% random sampling, a total of 233 articles were included. Among them, 149 (63.9%) reported the study design in at least 1 section: the title (n = 64 [27.5%]), abstract (n = 92 [39.5%]), or methods (n = 129 [55.4%]). However, only 58 of 149 (38.9%) appropriately reported their study design. The reasons for inappropriate study design reporting were (1) mislabeling of nondesign elements as study designs (eg, subgroup analysis, retrospective study [n= 37]), (2) incorrect classification of study design (eg, describing a cohort study as a case-control study [n = 36]), and (3) insufficient detail in describing the study design (eg, labeling a nonrandomized study as a phase trial, labeling a cohort study as a longitudinal study [n = 18]). Reporting guidelines in submission policies, including reference to the EQUATOR Network, showed no significant association with either the reporting or appropriateness of study designs.

Conclusions More than half of the articles in Korean medical journals reported their study designs, yet only small fractions were deemed appropriate. Enhanced education for researchers on study design is necessary to improve reporting quality.

References

1. Seo HJ, Kim SY, Lee YJ, et al. A newly developed tool for classifying study designs in systematic reviews of interventions and exposures showed substantial reliability and validity. *J Clin Epidemiol*. 2016;70:200-205. doi:10.1016/j.jclinepi.2015.09.013

2. The EQUATOR Network. Accessed January 13, 2025. https://www.equator-network.org

[1]Department of Family Medicine, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Korea, hallymfm@gmail.com; [2]College of Nursing, Yonsei University, Seoul, South Korea; [3]Medical Library, Hanyang University Guri Hospital, Guri, Korea.

Conflict of Interest Disclosures None reported.

Comparative Analysis of Expert, Clinician, and Consumer Interactions With Summary of Findings Tables: A Quasi-Experimental Study

Nina Vitlov,[1] Nensi Bralić,[1] Tina Poklepović Peričić,[2] Daniel Garcia-Costa,[3] Emilia López-Iñesta,[4] Elena Álvarez-García,[3] Francisco Grimaldo,[3] Ana Marušić[1]

Objective To explore how GRADE and Cochrane experts, clinicians, and health care consumers interact with Summary of Findings (SoF) tables from systematic reviews of evidence when answering questions about table content.

Design A quasi-experimental study was conducted from February to June 2025, using the Read&Learn platform, an online tool designed for monitoring the interaction with a word problem or a table.[1,2] The sample size was calculated at 25 participants per group to detect a mean (SD) of 4 (5) points' difference in the number of correct answers (range, 0-16), with 80% power and 5% α level. The study intervention was blurring of SoF table cells, which the participants were required to navigate and open in order to answer questions about 4 different SoF tables, alternating between the question screen and table cells. All necessary data for answering the questions were contained within the target cells, requiring no

additional calculations. We measured the number of correct answers to the questions about SoF tables, total number of table cells visited, number of target table cells visited, number of nontarget table cells visited, total time spent on the quiz from beginning to end, time spent on initial reading, total time spent reading table cells, time spent on reading target table cells, and time spent on reading nontarget table cells. The study was approved by the Ethics Committee of the University of Split School of Medicine, and all participants gave informed consent.

Results We collected the data from 40 experts, 40 clinicians, and 40 health care consumers (**Table 25-0884**). There were no significant differences between experts, clinicians, and health care consumers in terms of quiz performance, with all groups achieving a similar number of correct answers. Additionally, total time spent on the quiz, initial reading, and quiz questions did not differ between the groups. Significant differences were observed in their interaction with the SoF tables. Experts visited fewer numbers of table cells in total and fewer nontarget segments and spent the least amount of time reading the nontarget segments, compared with clinicians and health care consumers. They also spent the least amount of time reading the question statement.

Conclusions Experts in evidence synthesis, Cochrane systematic review authors, and GRADE experts showed different interaction patterns with the SoF table, possibly reflecting their greater familiarity with the content. Future studies, including qualitative approaches, should further explore how different users approach and use the information presented in SoF tables.

Table 25-0884. Summary Statistics for 13 Measures of the Interaction With Summary of Findings (SoF) Tables by Experts, Clinicians, and Health Care Consumers[a]

Parameter	Median (95% CI)			*P* value[b]
	Experts (n = 40)	Clinicians (n = 40)	Health care consumers (n = 40)	
Correct answers, No. (maximum, 16 points)	14.0 (14.0-15.0)	15.0 (14.0-16.0)	14.0 (13.0-14.7)	.07
Total time spent on quiz from beginning to end, s	337.5 (208.0-489.3)	591.3 (39.1-728.5)	410.3 (334.3-629.3)	.14
Time spent on initial reading, s[c]	0 (0-6.6)	0 (0-7.3)	4.3 (0-22.9)	.44
Time spent on quiz questions, s	635.2 (487.6-906.63)	1061.3 (833.7-1392.8)	867.9 (580.0-1120.6)	.12
Segments visited, total No.	71.5 (63.0-87.7)[d]	88.0 (77.0-118.3)	90.5 (80.4-109.9)	.04
Total time spent reading segments during search, s	171.5 (109.4-219.3)	170.2 (114.9-261.9)	147.5 (117.2-239.4)	.81
Total time spent on reading target segments, s	120.5 (86.6-155.1)	138.7 (85.3-188.7)	118.1 (82.7-157.8)	.71
Target segments visited, No.	44.5 (40.3-49.7)	51.5 (44.7-63.0)	46.0 (39.7-55.0)	.26
Total time spent on reading nontarget segments, s	44.8 (35.1-75.2)[d]	84.3 (63.8-126.4)	91.3 (56.1-145.5)	.02
Nontarget segments visited, No.	27.5 (21.7-35.7)[d]	37.5 (28.7-57.7)	42.5 (33.0-55.7)	.004
Time spent reading the question statement, s	160.8 (124.3-196.5)[d]	249.7 (188.1-311.5)	253.3 (171.9-355.1)	.004
No. of times reading the question statement	34.5 (26.4-39.0)	37.5 (33.3-39.0)	33.5 (28.0-36.7)	.60
No. of times searching in text for given questions	23.0 (21.3-26.7)	23.0 (22.0-26.0)	23.5 (21.0-27.0)	.87

[a]For each Read&Learn parameter, summary data are presented for all 4 SoF tables.

[b]Calculated with Kruskal-Wallis test.

[c]Time spent on initial reading of the SoF table after the start of the task before switching to the questions.

[d]Significantly different from other groups ($P < .05$) based on Kruskal-Wallis test.

References

1. Matas J, Tokalić R, García-Costa D, et al. Tool to assess recognition and understanding of elements in Summary of Findings table for health evidence synthesis: a cross-sectional study. *Sci Rep*. 2023;13:18044. doi:10.1038/s41598-023-45359-x

2. Sanz MT, López-Iñesta E, García-Costa D, Grimaldo F. Measuring arithmetic word problem complexity through reading comprehension and learning analytics. *Mathematics*. 2020;8(9):1556. doi:10.3390/math8091556

[1]Department of Research in Biomedicine and Health, Center for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia, nina.vitlov@mefst.hr; [2]Department of Prosthodontics, School of Dental Medicine, University of Split School of Medicine, Split, Croatia; [3]Department of Computer Science, Universitat de València, València, Spain; [4]Department of Didactics of Mathematics, Universitat de València, València, Spain.

Conflict of Interest Disclosures Nina Vitlov is funded by the Croatian Science Foundation under the Programme for Career Development of Early Career Researchers-Training of New Doctoral Students, NPOO (C3 2 R2-I1). Ana Marušić is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Replication and Impact of Positive Secondary Findings in Negative or Neutral Cardiovascular Trials

Sina Rashedi,[1] Farbod Zahedi Tajrishi,[2] Ashkan Hashemi,[3] Isaac Dreyfus,[4] Nicholas Varunok,[5] John Burton,[6] Seng Chan You,[7] Bjorn Redfors,[8,9] Gregory Piazza,[1,10] Joshua D. Wallach,[11] Lesley Curtis,[12] Sanjay Kaul,[13] David J. Cohen,[14,15] Roxana Mehran,[16] Mitchell S. V. Elkind,[17] Flavia Geraldes,[18] Joseph S. Ross,[19,20] Jane A. Leopold,[10] Harlan M. Krumholz,[19,21] Gregg W. Stone,[16] Behnood Bikdeli[1,10,19]

Objective A large number of cardiovascular randomized clinical trials (RCTs) do not meet their primary outcome (neutral or negative trials). Some of these negative/neutral RCTs show positive secondary findings. We aimed to assess the proportion of cardiovascular disease– or stroke-related RCTs with negative or neutral primary results but positive

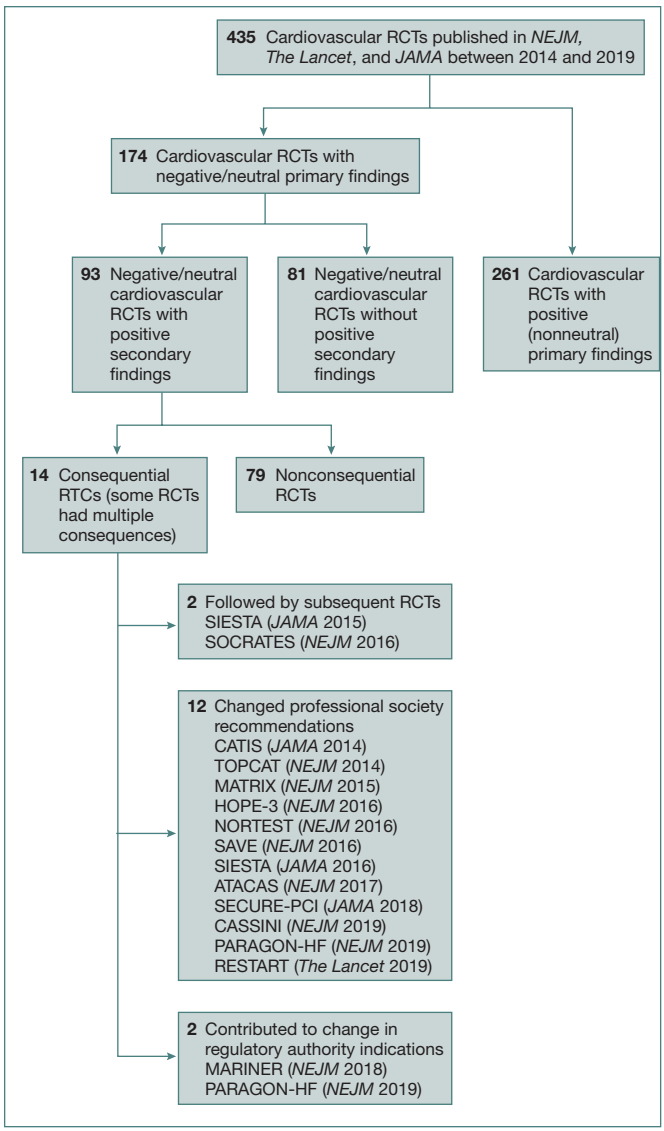
secondary findings among all negative/neutral cardiovascular RCTs published in the highest-impact medical journals and to evaluate whether these findings were pursued in subsequent confirmatory RCTs or influenced clinical practice recommendations or regulatory decisions.

Design We searched PubMed to identify cardiovascular RCTs published in 3 major clinical journals (*NEJM*, *The Lancet*, and *JAMA*) between 2014 and 2019. Negative/neutral RCTs with positive secondary findings were defined as those that did not meet their (co)primary outcome(s) ($P > .05$) but showed positive effects in prespecified subgroup analyses for the primary outcome and/or for prespecified secondary outcomes among the entire study population. Secondary outcomes and subgroup analyses were considered if they were reported as prespecified in the trial publications. We investigated whether these positive secondary findings were examined in subsequent confirmatory RCTs via PubMed searches or directly resulted in changes in either professional society recommendations or regulatory authority endorsements in the US or Europe through December 2024. Two investigators independently performed the literature search and data extraction, with discrepancies resolved by a third investigator.

Results Overall, 435 cardiovascular RCTs were identified; the median time from the index trial to December 2024 was 7.5 (IQR, 5.5-9.5) years. Of these 435 RCTs, the primary outcome(s) were not met in 174 trials (40.0%), of which 93 (53.4%) had at least 1 positive secondary finding. Among these trials, 51 (54.8%) had only positive secondary outcomes, 20 (21.5%) had only positive findings in subgroup analyses, and 22 (23.7%) had positive findings in both secondary outcomes and subgroup analyses. Among the 93 negative/neutral RCTs with positive secondary findings, 14 (15.1%) were consequential: subsequent RCTs were performed due to positive secondary findings for 2 trials; the secondary results from 12 RCTs contributed to changes in professional society recommendations; and 2 trials contributed to changes in labeled indications (**Figure 25-0928**).

Conclusions Although nearly half of the negative/neutral cardiovascular trials published in the highest-impact journals report positive secondary findings, only a small proportion directly influence practice or policy, and very few are followed by subsequent confirmatory RCTs. Our findings caution against overemphasizing positive secondary results from cardiovascular RCTs and underscore the need for subsequent confirmatory RCTs to evaluate positive secondary findings from otherwise negative/neutral trials. Further research is planned to assess the generalizability of these findings in major specialty cardiovascular journals and to identify trial- and outcome-level characteristics that influence the selection of positive secondary findings for replication in subsequent trials.

Figure 25-0928. Negative or Neutral Cardiovascular Randomized Clinical Trials (RCTs) With Positive Secondary Findings and Their Subsequent Impact



[1]Thrombosis Research Group, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US, srashedi@bwh.harvard.edu; [2]Tulane University School of Medicine, New Orleans, LA, US; [3]Program for the Care and Study of the Aging Heart, Department of Medicine, Weill Cornell Medicine, New York, NY, US; [4]Division of Cardiology, Department of Medicine, University of California Los Angeles, Los Angeles, CA, US; [5]Vanderbilt University Medical Center, Nashville, TN, US; [6]University of Southern California, Los Angeles, CA, US; [7]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea; [8]Department of Cardiology, Sahlgrenska University Hospital, Gothenburg, Sweden; [9]Cardiovascular Research Foundation, New York, NY, US; [10]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US; [11]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, US; [12]Duke University School of Medicine, Department of Population Health Sciences, Durham, NC, US; [13]Department of Cardiology, Cedars-Sinai Medical Center, Los Angeles, CA, US; [14]Clinical Trials Center, Cardiovascular Research Foundation, New York, NY, US; [15]St Francis Hospital and Heart Center, Roslyn, NY, US; [16]Icahn School of Medicine at Mount Sinai, New York, NY, US; [17]Department of Neurology, Columbia University Irving Medical Center, New York, NY, US; [18]*The Lancet*, London, United Kingdom; [19]YNHH/Yale Center for Outcomes Research and Evaluation (CORE), New Haven, CT, US; [20]Department of General Internal Medicine, Yale School of Medicine, New Haven, CT, US; [21]Section of Cardiovascular Medicine, Yale School of Medicine, New Haven, CT, US.

Conflict of Interest Disclosures Outside the submitted work, Behnood Bikdeli reported receiving the following support: a Career Development Award from the American Heart Association and VIVA Physicians (#938814), the Scott Schoen and Nancy Adams IGNITE Award, the Mary Ann Tynan Research Scientist Award from the Mary Horrigan Connors Center for Women's Health and Gender Biology at Brigham and Women's Hospital, and the Heart and Vascular Center Junior Faculty Award from Brigham and Women's Hospital; serving as consulting expert on behalf of the plaintiff for litigation related to 2 specific brand models of IVC filters (although he has not been involved in the litigation in 2022-2025, nor has he received any compensation in 2022-2025); serving as a member of the Medical Advisory Board for the North American Thrombosis Forum (now VascuLearn Network) and on the Data Safety and Monitoring Board of the NAIL-IT trial funded by the National Heart, Lung, and Blood Institute and Translational Sciences; serving as a collaborating consultant with the International Consulting Associates and the US Food and Drug Administration in a study to generate knowledge about the utilization, predictors, retrieval, and safety of IVC filters; receiving compensation as an associate editor for *NEJM Journal Watch Cardiology*, as an associate editor for *Thrombosis Research*, and as an executive associate editor for *JACC*; and serving as a section editor for *Thrombosis and Haemostasis* without compensation.

Additional Information Behnood Bikdeli is a co–corresponding author (behnood.bikdeli@yale.edu).

Public Availability of Randomized Clinical Trial Protocols: A Repeated Cross-Sectional Study

Christof Manuel Schönenberger,[1] Malena Chiaborelli,[1] Ala Taji Heravi,[1] Lukas Kübler,[2] Pooja Gandhi,[3] Zsuzsanna Kontar,[4] Julia Hüllstrung,[1] Mona Elalfy,[1] Jan Glasstetter,[1] Dmitry Gryaznov,[5] Belinda von Niederhäusern,[6] Anette Blümle,[7] Jason W. Busse,[8] Szimonetta Lohner,[4] Sally Hopewell,[9] Alexandra Griessbach,[1] Matthias Briel,[1,10] Benjamin Speich[1]

Objective Making protocols of randomized clinical trials (RCTs) publicly available is important for the trustworthiness and quality of medical research.[1] In a previous study, only 36% of RCTs that received ethical approval in 2012 had a publicly available protocol.[2] Repeating this study with RCTs approved in 2016, we aimed to generate current evidence on the availability of RCT protocols, sources where they are shared, and changes over time.

Design Using a sample of RCTs receiving ethical approval in 2016 in Switzerland, Canada, Germany, or the UK,[3] we investigated the number of available protocols. For all RCTs, we checked if a result publication was available by searching PubMed, Google Scholar, trial registries, and Google. We extracted baseline characteristics for all RCTs. Up to June 2024, we systematically searched for (1) protocols available as peer-reviewed publications, (2) protocols attached to trial registries, and (3) protocols shared with result publications of RCTs. Multiple sources per protocol were possible. All searches and data extraction were performed in duplicate. We used multivariable logistic regression to examine the association of protocol availability with trial characteristics such as sample size, drug vs nondrug interventions, multicenter vs single-center status, and RCT approval in 2016 vs 2012.

Results Of 347 included RCTs, 228 (65.7%) had an available protocol (**Table 25-0957**); 150 protocols (43.2%) were available as files on trial registries, 91 (26.2%) as supplementary material to a result publication, and 81 (23.3%) as peer-reviewed publications. Protocol availability improved over time (65.7% in 2016 vs 36.2% in 2012), particularly in industry trials (83.4% in 2016 vs 34.6% in

Table 25-0957. Characteristics of Included RCTs and Availability of Protocols[a]

Characteristic	Year of ethical approval					
	2012			2016		
	All included RCTs (n = 326)	RCTs with publicly available protocol (n = 118)	RCTs without publicly available protocol (n = 208)	All included RCTs (n = 347)	RCTs with publicly available protocol (n = 228)	RCTs without publicly available protocol (n = 119)
Sponsorship						
Industry	179 (54.9)	62/179 (34.6)	117/179 (65.4)	181 (52.2)	151/181 (83.4)	30/181 (16.6)
Nonindustry	147 (45.1)	56/147 (38.1)	91/147 (61.9)	166 (47.8)	77/166 (46.4)	89/166 (53.6)
Drug vs nondrug intervention						
Drug	207 (63.5)	77/207 (37.2)	130/207 (62.8)	212 (61.1)	163/212 (76.9)	49/212 (23.1)
Nondrug	119 (36.5)	41/119 (34.4)	78/119 (65.5)	135 (38.9)	65/135 (48.2)	70/135 (51.8)
Single center vs multicenter						
Single center	60 (18.4)	12/60 (20.0)	48/60 (80.0)	82 (23.6)	22/82 (26.8)	60/82 (73.2)
Multicenter	266 (81.6)	106/266 (39.8)	160/266 (60.2)	265 (76.4)	206/265 (77.7)	59/265 (22.3)
International	199 (74.8)	79/199 (39.7)	120/199 (60.3)	207 (59.6)	170/207 (82.1)	37/207 (17.9)
No. of participants						
<100	73 (22.4)	14/73 (19.2)	59/73 (80.8)	79 (22.8)	29/79 (36.7)	50/79 (63.3)
100-500	151 (46.3)	45/151 (29.8)	106/151 (70.2)	191 (55.0)	131/191 (68.6)	60/191 (31.4)
>500	102 (31.3)	59/102 (57.8)	43/102 (42.2)	77 (22.2)	68/77 (88.3)	9/77 (11.7)

Abbreviation: RCT, randomized clinical trial.

[a]Data are presented as No. (%) or No./total No. (%).

2012), but improved little in nonindustry trials (46.4% 2016 vs 38.1% 2012). In a regression model with pooled data from trials approved in 2012 and approved in 2016, trials with a medium sample size (100-500 participants) (odds ratio [OR], 2.12; 95% CI, 1.32-3.43; $P = .001$), trials with a large sample size (>500 participants) (OR, 6.12; 95% CI, 3.45-11.07; $P < .001$), and multicenter trials (vs single-center trials) (OR, 2.66; 95% CI, 1.59-4.52; $P < .001$) showed significantly higher odds of having an available protocol. A significant interaction term indicated dependence of sponsorship and year (OR, 0.13; 95% CI 0.06-0.27; $P < .001$).

Conclusions The availability of protocols increased in RCTs approved in 2016 compared with RCTs approved in 2012. This was mainly driven by industry-sponsored trials and may be explained by updated US regulations. Since 2017, all trials investigating drugs or devices that are or are intended to be approved, licensed, or cleared by the US Food and Drug Administration must have a publicly available protocol. Efforts to further improve protocol availability should be continued, especially in nonindustry-sponsored RCTs.

References

1. Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. 2014;383(9913):257-266. doi:10.1016/S0140-6736(13)62296-5

2. Schönenberger CM, Griessbach A, Taji Heravi A, et al. A meta-research study of randomized controlled trials found infrequent and delayed availability of protocols. *J Clin Epidemiol*. 2022;149:45-52. doi:10.1016/j.jclinepi.2022.05.014

3. Gryaznov D, Odutayo A, von Niederhäusern B, et al. Rationale and design of repeated cross-sectional studies to evaluate the reporting quality of trial protocols: the Adherence to Spirit Recommendations (ASPIRE) study and associated projects. *Trials*. 2020;21(1):896. doi:10.1186/s13063-020-04808-y

[1]Division of Clinical Epidemiology, Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland, christofmanuel.schoenenberger@usb.ch; [2]Department of Biomedicine, University Hospital Basel, University of Basel, Basel, Switzerland; [3]University Health Network, Toronto, Ontario, Canada; [4]MTA-PTE Lendület "Momentum" Evidence in Medicine Research Group, Department of Public Health Medicine, Medical School, University of Pécs, Pécs, Hungary; [5]Viatris Innovation GMBH, Allschwil, Switzerland; [6]Roche Pharma AG, Grenzach-Wyhlen, Germany; [7]Clinical Trials Unit, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany; [8]Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada; [9]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK; [10]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada.

Conflict of Interest Disclosures None reported.

Funding/Support This project was supported by the Swiss Federal Office of Public Health (to Matthias Briel). Christof Manuel Schönenberger was funded by the Janggen Pöhn Foundation and the Swiss National Science Foundation (MD-PhD grant number 323530_221860).

Role of the Funder/Sponsor The funders had no role in the conduct of the study, the analysis of the data, or the writing of the abstract.

Acknowledgment We thank all participating research ethics committees from Germany (Freiburg), Switzerland (Basel, Bellinzona, Bern, Geneva, Lausanne, St Gallen, Frauenfeld, and Zurich), Canada (Hamilton), and the UK (National Health Service Health Research Authority) for their support and cooperation.

Adherence to the WHO Statement on Public Disclosure of Clinical Trial Results by Trials Published in High Impact Factor Journals

Carolina Grana,[1,2,3] Lina Ghosn,[1,2,3] Carolina Riveros,[1,2,3] Philippe Ravaud,[1,2,3] Isabelle Boutron[1,2,3]

Objective The World Health Organization (WHO) requests that "key outcomes are to be made publicly available within 12 months of study completion by posting to the results section of the primary clinical trial registry."[1] In some regions, this requirement is also mandated by law[2]; however, evidence indicates low adherence.[3] The extent to which journal editors support this request remains unclear. This study aimed to assess adherence to the WHO request among randomized clinical trials (RCTs) published in high-impact journals and to examine whether journal editors encourage this practice in their recommendations to authors.

Design We conducted a cross-sectional study and reported our findings according to the STROBE guideline. On April 18, 2025, we searched PubMed using the terms *date-create: 01/01/2024 to 01/01/2025 AND journal: [Name of the journal]*, applying the "randomized controlled trial" filter. We selected the 5 journals with the highest impact factors in the "medicine, general & internal" category and in selected medical specialties based on the 2023 Clarivate Journal Citation Reports (Web of Science). We included RCTs evaluating pharmacologic, biologic, or medical device interventions. Data were extracted in duplicate from each trial publication and corresponding primary registry entry. We recorded the publication date, the trial's primary completion date, and results status in the registry. The primary outcome was the proportion of RCTs that at the time of publication had results posted in the primary registry within 12 months of trial completion. The analysis was primarily descriptive, comparing the timing of trial completion, results posting, and publication.

Results Preliminary results (publications up to June 2024) identified 185 RCTs published across 18 journals. Among these, 142 (77%) evaluated pharmacologic interventions and 108 (58%) were multinational. Trial sites included at least 1 European Union country in 102 trials (55%), the US in 94 (51%), and the UK in 74 (40%). All 185 trials were registered, and 163 (88%) registrations were prospective. A total of 100 trials (54%) were published more than 12 months after their primary completion date. Among these, at the time of publication, only 23 (23%) had adhered to WHO

requirements by having results posted in the trial registry within 12 months of completion. Conversely, 83 trials (45%) were published within 12 months of completion, and 45 of these (54%) had results posted in the registry when the trial report was published. Completion date was not reported for 2 trials (1%). None of the 18 journals assessed requested public posting of trial results in registries.

Conclusions Preliminary findings from trials included in this study, all published in high–impact factor journals, revealed low adherence to the WHO call for timely posting of results to the registries.

References

1. WHO statement on public disclosure of clinical trial results. World Health Organization. Accessed January 10, 2025. https://www.who.int/news/item/09-04-2015-japan-primary-registries-network

2. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. *N Engl J Med*. 2011;364(9):852-860. doi:10.1056/NEJMsa1012065

3. DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *Lancet*. 2020;395(10221):361-369. doi:10.1016/S0140-6736(19)33220-9

[1]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAe, Centre for Research in Epidemiology and Statistics (CRESS), Paris, France, isabelle.boutron@aphp.fr; [2]Centre d'Épidémiologie Clinique, AP-HP, Hôpital Hôtel-Dieu, Paris, France; [3]Cochrane France, Paris, France.

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.

Additional Information Carolina Grana (carolinaestela.granapossamai@aphp.fr) and Lina Ghosn (lina.elchall@aphp.fr) are co–corresponding authors.

Prevalence of Prospective Registration and Primary Outcome Discrepancies in Recently Published Randomized Controlled Trials

Ioana Alina Cristea,[1,2] Florian Naudet,[2,3,4] Guillaume Cabanac,[4,5] John P. A. Ioannidis[2,6,7,8]

Objective An older, field-wide analysis[1] of studies published between 2005 and 2017 estimated the prevalence of prospectively registered randomized controlled trials (RCTs) at 21%. Recent estimates indicated higher rates, eg, approximately 59% for rheumatology RCTs published between 2009 and 2022 in 5 International Committee of Medical Journal Editors (ICMJE) journals.[2] There are no recent field-wide estimates of the prevalence of prospective registration across clinical specialties and journals and of the related outcome reporting bias (ie, discrepancies between registered and published outcomes). We report these across a randomly selected sample of recently published RCTs.

Design We conducted a retrospective cohort study, reported following the STROBE guidelines. We searched PubMed with the Cochrane Highly Sensitive Search Strategy[3] on January 15, 2025. RCTs evaluating health-related interventions and outcomes (per the ICMJE definition) published in 2024 were included. Search results were randomized using Entrez Programming Utilities and the Linux shuf command, and a sample of 1720 records was selected. One researcher (I.A.C.) manually checked titles and abstracts to select RCTs and extracted trial registration information from publications reporting on primary outcomes (POs). Prospective registration was defined as submission date to registry preceding recruitment or study start or postdating by less than 30 days. We favored actual vs estimated study start dates when available in publications or registries. For prospectively registered trials, 1 researcher (I.A.C.) extracted all POs registered and declared in publications, and tabulated all discrepancies regarding outcome domain, measurement, or time point.

Results Of 324,177 records identified, 1720 were screened, leading to 136 eligible RCTs. Six trials were excluded (inaccessible full text [n = 3]; Chinese language [n = 3]). Eighty-two of 130 publications (63%) reported study start dates, with a median (IQR) of 2020 (2) (range, 2013-2023). Sixty-nine of 130 RCTs (53% [95% CI, 44%-62%]) were prospectively registered (67 in ICMJE-compliant registries), including 4 registered while recruiting. Sixty-one of 130 RCTs (47% [95% CI, 38%-56%]) were declared nonregistered (n = 4), were registered retrospectively (n = 27), indicated a registration number for another trial (n = 2), or did not mention registration (n = 28). Registration delays were a median (IQR) of 632 (533) (range, 49-1524) days. For 24 of 60 nonprospectively registered trials, journal instructions to authors explicitly required prospective registration or compliance with ICMJE or Declaration of Helsinki registration requirements. Of 69 prospectively registered trials, 6 did not declare POs in publications and 2 insufficiently specified POs in the registries, while 37 (54% [95% CI, 41%-66%]) had no substantive discrepancies between registry and publication. The remaining 24 of 69 RCTs (35% [95% CI, 24%-47%]) included 1 (n = 17), 2 (n = 6), or 3 (n = 1) types of discrepancies (**Table 25-1102**).

Conclusions In a random sample of RCTs published in 2024 across clinical specialties, approximately half were prospectively registered and approximately one-third of these contained substantial changes between registered and reported primary outcomes.

References

1. 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

2. Mongin D, Buitrago-Garcia D, Capderou S, et al. Prospective registration of trials: where we are, why, and how we could get better. *J Clin Epidemiol*. 2024;176. doi:10.1016/j.jclinepi.2024.111586

Table 25-1102. Primary Outcome (PO) Discrepancies Between Registry vs Publications for Prospectively Registered Trials (N = 69)

Discrepancy definition	Registered PO not reported in publication	Reported PO in publication not registered	Registered PO reported as secondary (downgrading)	Reported PO in publication registered as secondary (upgrading)	Same PO outcome domain but different assessment method in registry vs publication	Changes in PO assessment time frame in registry vs publication[a]
RCTs with discrepancy, No.[b]	4	2	9	3	6	8
Proportion, % (95% CI)[c]	6 (2-14)	3 (0.3-10)	13 (6-23)	4 (0.9-12)	9 (3-18)	12 (5-22)

Abbreviation: RCT, randomized controlled trial.

[a]Registered time points nonreported, downgraded from primary, nonregistered time points added in publication, upgraded from secondary.

[b]Some RCTs contained more than 1 discrepancy.

[c]Of the total number of prospectively registered RCTs.

3. Lefebvre C, Glanville J, Briscoe S, et al. Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions, Version 6.5*. Cochrane; 2024

[1]Department of General Psychology, University of Padova, Padova, Italy, ioanaalina.cristea@unipd.it; [2]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, US; [3]Université Rennes, CHU Rennes, Inserm, Centre d'investigation clinique de Rennes (CIC1414), service de pharmacologie clinique, Institut de recherche en santé, environnement et travail (Irset), UMR S 1085, EHESP, 35000, Rennes, France; [4]Institut Universitaire de France (IUF), Paris, France; [5]Université de Toulouse, IRIT (UMR 5505 CNRS), Toulouse, France; [6]Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, US; [7]Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, US; [8]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, US.

Conflict of Interest Disclosures John P. A. Ioannidis is a member of the Peer Review Congress Advisory Board but was not involved in the review or decision for this abstract.

Funding/Support Ioana Alina Cristea is supported by a European Research Council (ERC) Starting Grant (grant agreement: 101042701; https://cordis.europa.eu/project/id/101042701), funded by the European Union (EU). Florian Naudet received funding from the French National Research Agency, the French Ministry of Health, and the French Ministry of Research and is a work package leader in the OSIRIS project (Open Science to Increase Reproducibility in Science) and for the doctoral network MSCA-DN SHARE-CTD (HORIZON-MSCA-2022-DN-01 101120360), funded by the EU. Guillaume Cabanac received funding from the Institut Universitaire de France. John P. A. Ioannidis is supported by an unrestricted gift from Sue and Bob O'Donnell to Stanford University.

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 abstract.

Additional Information John P. A. Ioannidis is a co–corresponding author (jioannid@stanford.edu).

Virtual

Developing a Harmonized Approach for Reporting Irradiation Protocols and Methods for Research Using X-ray Irradiators

Warren Stern,[1] Ioanna Iliopulos,[2] Christopher Boyd[2]

Objective The lack of agreement on minimum standards for disclosure of irradiation parameters results in the need to repeat costly studies, uncertainty in observed effects, and difficulty validating research outcomes. In turn, the Brookhaven National Laboratory (BNL) conducted a meta-study to gain greater quantitative and qualitative insights into this issue. After the results were published in 2022,[1,2] 2 annual Compatibility in Irradiation Research Protocols Expert Roundtable (CIRPER) meetings were conducted by BNL with the support of the National Nuclear Security Administration's Office of Radiological Security. We describe the process of advocacy research that was implemented to address this issue—that is, conducting a technical meta-analysis, identifying gaps and actions based on its results, sufficiently addressing those gaps, and then publishing the outcome along with the recommendation in a peer-reviewed article to encourage dissemination and adoption by the wider community.

Design The study that BNL published in 2022 identified over 450 published articles and demonstrated that although source specifications (eg, model, energy, and spectra) are well reported, experimental methodologies and irradiation protocols are not. Based on these findings, participants for CIRPER were carefully selected to represent a diverse segment of the radiation research community, such as from the fields of medical physics and radiobiology, working in academic, health care, and industry settings alongside federal partners, national laboratory personnel, and device manufacturer representatives. The meeting provided a venue to bring together diverse stakeholders that typically do not collectively interact and consisted of detailed workshops designed to parse out necessary methodological parameters for published and federally funded research.

Results To increase transparency and reproducibility, experts at the meeting agreed on a standard set of disclosure requirements researchers working with x-ray irradiators should include when publishing their work (**Table 25-0890**). The meeting also resulted in further discussions on how session participants can support CIRPER's efforts to harmonize the reporting of irradiation protocols and methods.

Conclusions As the research community moves to x-ray–based irradiation technologies from cesium 137–based ones, reporting parameters must adapt and update accordingly. The BNL meta-analyses uncovered data that quantifies this

Table 25-0890. Summary of Recommendations Regarding Methodological Disclosure Requirements Based on the Compatibility in Irradiation Research Protocols Expert Roundtable (CIRPER) Consensus

Recommended methodological disclosure requirements
Device parameters
A. Manufacturer and model
B. Energy (kVp)
C. Time (mAs) or dose rate (Gy/min)
D. Air kerma or dose output (Gy/min) as provided by the manufacturer's datasheets (Gy/min)
E. Absorbed dose to target (Gy)
F. Filters (materials and thickness) and half-value layer (HVL) whenever available
Machine calibration details
A. Calibration detector (manufacturer and model)
B. The energy at which the dosimeter was calibrated, and the dose quantity used for calibration (Dw or air kerma)
C. Frequency and date
Experimental setup
A. Field size
B. Surface of sample distance from source and orientation
C. Sample description
D. Sample shielding
E. Sample holder
F. Other sources of scatter
G. Photograph of the experimental setup (optional)

Abbreviation: kerma, kinetic energy released per unit mass.

barrier. A strategy for advocacy research was designed and implemented accordingly to appropriately address this problem by (1) further identifying gaps in the reporting process, (2) gathering relevant stakeholders to design an informed approach, and (3) gaining commitment from journal editorial boards, funding agencies, and researchers to adopt CIRPER recommendations. This effort has already had a positive impact on the quality of peer-reviewed and funded research and serves as a model for improving other aspects of peer review and publication.[3]

References

1. Stern W, Zia S, Boyd C, Iliopulos I, Peiris P, Fedurin M. *Alternative Technologies Meta Study Report*. US Department of Energy; 2022. doi:10.2172/1895064

2. Stern W, Zia S, Boyd C, Iliopulos I. *Alternative Technologies Adopters Outreach Survey Report*. US Department of Energy; 2022. doi:10.2172/1895065

3. Stern W, Alaei P, Berbeco R, et al. Achieving consistent reporting of radiation dosimetry by adoption of compatibility in irradiation research protocols expert roundtable (CIRPER) recommendations. Radiat Res. 2024;201(3):267-269. doi:10.1667/RADE-23-00234.1

[1]Brookhaven National Laboratory, Upton, NY, US, wstern@bnl.gov; [2]MARC Strategics, Ann Arbor, MI, US.

Conflict of Interest Disclosures None reported.

Funding/Support This work was funded and sponsored by an agency of the US government.

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 manuscript; and decision to submit the manuscript for publication.

Reporting Guidelines

In-person

Challenges in Achieving Uptake and Journal Endorsement of the ACcurate COnsensus Reporting Document Guideline: An Observational Study

Christopher C. Winchester,[1] Mark J. Rolfe,[1] William T. Gattrell,[2] Patricia Logullo,[3] Keith Goldman,[4] Amy Price,[5] Paul Blazey,[6] Esther J. van Zuuren,[7] Niall Harrison[8]

Objective The ACcurate COnsensus Reporting Document (ACCORD) guideline supports the reporting of biomedical studies involving any consensus method.[1] We evaluated its uptake by bibliometric analysis and qualitative and quantitative assessment of implementation activities 1 year post publication.

Design An implementation plan[2] was established by the ACCORD Steering Committee before guideline publication, including developing an explanation and elaboration (E&E) document, writing journal editorials, presenting at congresses, and encouraging journals and publishers to include ACCORD in author instructions. Guideline views and downloads (via *PLOS Medicine*,[1] the publishing journal) and citations (Dimensions, plus additional references from the Steering Committee's reference libraries) were obtained for the period from January 23, 2024, to January 22, 2025. Likely journal and publisher beneficiaries of ACCORD were identified by a literature search to determine the 10 journals that published the most consensus studies (Web of Science for 2020-2023) and other publications at the discretion of the Steering Committee. Relevant journal and publisher stakeholders were contacted directly or indirectly (eg, via journal websites). Activities were evaluated against the implementation plan[2]; journal and publisher responses were described.

Results In the 1 year following publication, the ACCORD guideline was viewed 13,752 times, downloaded (full-text PDF) 3790 times, and cited 105 times (5 citations [4.8%] were in articles by ACCORD authors, including the ACCORD

E&E[3]). Sources of guideline citations included consensus studies (67.6% [71 of 105]), reviews (11.4% [12 of 105]), study protocols (8.6% [9 of 105]), other reporting standards and guidelines (6.7% [7 of 105]), and editorials (3.8% [4 of 105]). Of the journals contacted by ACCORD Steering Committee members (between August 24, 2023 [preprint publication], and January 22, 2025), *BMJ Open*, *British Journal of Dermatology*, and *Journal of Clinical Epidemiology* mandated the use of ACCORD for consensus studies, while the *British Journal of Sports Medicine*, *Internal Medicine Journal*, and *Pragmatic and Observational Research* encouraged its use. *PLOS* has confirmed its intention to include the guideline in their author instructions. Although a further 4 journals have expressed interest in ACCORD and 1 publisher has offered to raise awareness via a webinar, they are yet to commit to its inclusion in author instructions; 2 journals and publishers have not responded to inquiries.

Conclusions One year after publication, the ACCORD guideline has been cited and used to inform the design and reporting of consensus-based research. However, journal and publisher adoption has been limited, which may affect long-term uptake. Given the resources invested in developing reporting guidelines and their potential to improve reporting, the role of biomedical journals and publishers in their adoption warrants wider discussion.

References

1. Gattrell WT, Logullo P, van Zuuren EJ, et al. ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLOS Med*. 2024;21:e1004326. doi:10.1371/journal.pmed.1004326

2. Gattrell WT, Hungin AP, Price A, 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

3. Logullo P, van Zuuren EJ, Winchester CC, et al. ACcurate COnsensus Reporting Document (ACCORD) explanation and elaboration: guidance and examples to support reporting consensus methods. *PLoS Med*. 2024;21:e1004390. doi:10.1371/journal.pmed.1004390

[1]Oxford PharmaGenesis, Oxford, UK, chris.winchester@pharmagenesis.com; [2]Independent Medical Communications Professional, Oxfordshire, UK; [3]Bodleian Libraries, University of Oxford, Oxford, UK; [4]Medical Affairs + Health Impact, AbbVie, North Chicago, IL, US; [5]Dartmouth Institute for Health Policy & Clinical Practice (TDI), Geisel School of Medicine, Dartmouth College, Hanover, NH, US; [6]School of Kinesiology, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; [7]Leiden University Medical Centre, Leiden, the Netherlands; [8]OPEN Health Communications, London, UK.

Conflict of Interest Disclosures Christopher C. Winchester is an employee, director, and shareholder of Oxford PharmaGenesis; 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, University of Oxford. Mark J. Rolfe is an employee of Oxford PharmaGenesis. William T. Gattrell was an independent medical communications professional at the time of this study and is currently an employee of Bristol Myers Squibb. Keith Goldman is an employee of AbbVie. Niall Harrison is an employee of OPEN Health Communications.

Funding/Support Medical writing, editorial, and project management support for this abstract were provided by Oxford PharmaGenesis.

Acknowledgments Medical writing support was provided by Alison Chisholm, with editorial support from Jenny Thorp, and administrative support was provided by Mehraj Ahmed, Ryan Gamble, and Jessica Miller.

Additional Information No authors were reimbursed for participating in the initiative.

Transparent Reporting of Observational Studies Emulating a Target Trial: The TARGET Guideline

Aidan G. Cashin,[1,2] Harrison J. Hansford,[1,2] Miguel A. Hernán,[3,4,5] Sonja A. Swanson,[3,4,6] Hopin Lee,[7,8] Matthew D. Jones,[1,2] Issa J. Dahabreh,[3,4,5,9] Barbra A. Dickerman,[3,4] Matthias Egger,[10,11,12] Xabier Garcia-Albeniz,[3,13] Robert M. Golub,[14] Nazrul Islam,[15,16] Sara Lodi,[3,17] Margarita Moreno-Betancur,[18,19] Sallie-Anne Pearson,[20] Sebastian Schneeweiss,[21] Melissa K. Sharp,[22] Jonathan A. C. Sterne,[12,23,24] Elizabeth A. Stuart,[25] James H. McAuley[1,2]

Objective When randomized trials are unavailable or not feasible, data from observational studies can be used, if assumptions hold, to answer causal questions about the comparative effects of interventions by emulating a hypothetical pragmatic randomized trial (target trial). The reporting of studies that emulate a target trial is inconsistent and often incomplete, limiting transparency and reproducibility. To address this knowledge gap, we developed consensus-based guidance for reporting analyses of observational data that aim to estimate causal effects by explicitly emulating a target trial.

Design The TARGET (Transparent Reporting of Observational Studies Emulating a Target Trial)[1] guideline was developed using the Enhancing the Quality and Transparency of Health Research (EQUATOR) framework.[2] This included (1) a systematic review[3] of reporting practices in published studies that explicitly aimed to emulate a target trial, (2) a 2-round online survey (August 2023-March 2024; 18 expert participants from 6 countries) to generate agreement on the importance of candidate items selected from previous research and to identify additional items, (3) a 3-day expert consensus meeting (June 2024; 18 panelists) to refine the scope of the guideline and draft the checklist, and (4) an internal and external review and piloting activity with stakeholders and potential users (n = 108; September 2024-February 2025). The checklist was then refined based on feedback and finalized.

Results The 21-item TARGET checklist was organized into 6 sections (abstract, introduction, methods, results, discussion, other information) (**Table 25-0970**). The TARGET guideline was intended to be general, with a focus on nonrandomized studies of interventions explicitly emulating a parallel-group, individually randomized target trial, with adjustment for

baseline confounders. Key recommendations were to (1) summarize the causal question and the reason for emulating a target trial; (2) clearly specify the target trial protocol (ie, causal estimand, identifying assumptions, data analysis plan) and how these components were mapped to the observational data; and (3) for each causal estimand, report the estimate obtained and its precision, along with findings from additional analyses to assess the sensitivity of estimates to assumptions and design and analysis choices.

Table 25-0970. TARGET Checklist of Recommended Items to Address in Reports of Studies Emulating a Target Trial[a]

Item No.	Checklist Item				
Abstract					
1	a	Identify that the study attempts to emulate a target trial using observational data. State the study objectives and briefly summarize the specified target trial.			
	b	Report the data source(s) used for emulation.			
	c	Summarize key assumptions, statistical methods, findings, and conclusions.			
Introduction					
2	Background	Describe the scientific background of the study and the gap in knowledge.			
3	Causal question	Summarize the causal question.			
4	Rationale	Describe the rationale for emulating a target trial with the available data. Cite randomized trials informing the design of the target trial, if applicable.			
Methods					
5	Data source(s)	Cite the data source(s) contributing to the analyses and for each one, describe the following: original purpose, type, geographic location(s), setting, and time period. If relevant, describe how the data were linked or pooled.			
6	Target trial specification Specify the components of the target trial protocol that would answer the causal question.		7	Target trial emulation Describe how the components of the target trial protocol were emulated with the observational data, including how all variables were measured or ascertained.	
	Eligibility criteria			Eligibility criteria	
	a	Describe the eligibility criteria.		a	Describe how the eligibility criteria were operationalized with the data.
	Treatment strategies			Treatment strategies	
	b	Describe the treatment strategies that would be compared.		b	Describe how the treatment strategies were operationalized with the data.
	Assignment procedures			Assignment procedures	
	c	Report that eligible individuals would be randomly assigned to treatment strategies and may be aware of their treatment allocation.		c	Describe how assignment to treatment strategies was operationalized with the data.
	Follow-up			Follow-up	
	d	Clarify that follow-up would start at the time of assignment to treatment strategies. Specify when follow-up would end.		d	Clarify that follow-up would start at the time individuals were assigned to treatment strategies. Describe how the end of follow-up was operationalized with the data.
	Outcome(s)			Outcome(s)	
	e	Describe the outcome(s).		e	Describe how the outcome(s) were operationalized with the data.
	Causal contrast(s)			Causal contrast(s)	
	f	Describe the causal contrast(s) of interest, including effect measure(s).		f	Describe how the causal contrast(s) were operationalized with the data, including effect measure(s).
	Identifying assumptions			Identifying assumptions	
	g	Describe assumptions that would be made to identify each causal estimand. Describe the variables, if any, related to these assumptions.		g.i	For each causal estimand, describe assumptions made to identify it, including assumptions regarding baseline confounding due to lack of randomization.
				g.ii	Describe how the variables related to these assumptions were operationalized with the data
	Data analysis plan			Data analysis plan	
	h	For each causal estimand, describe the data analysis procedures and any associated statistical modeling assumptions, including approaches for handling missing data.		h.i	For each causal estimand, describe the data analysis procedures and any associated statistical modeling assumptions, including approaches for handling missing data.
				h.ii	For each causal estimand, describe any additional analyses conducted to assess the sensitivity of the results to the choice of operationalizations, assumptions, and analysis.
Results					
8	Participant selection	Report numbers of individuals assessed for eligibility, eligible, and assigned to each treatment strategy. A flow diagram is strongly recommended.			
9	Baseline data	Describe the distribution of characteristics of individuals at baseline, by treatment strategy.			
10	Follow-up	Summarize length of follow-up and describe reasons for end of follow-up for each treatment strategy and causal contrast.			
11	Missing data	Describe the frequency of missing data in all variables, by treatment strategy when applicable.			
12	Outcomes	Describe the frequency or distribution of each outcome, by treatment strategy.			
13	Effect estimates	Report the effect estimates for each causal contrast with corresponding measures of precision, including both absolute and relative measures of effect, when applicable.			
14	Additional analyses	Report results of all analyses to assess the sensitivity of the estimates to choices in operationalizations, assumptions, and analysis.			
Discussion					
15	Interpretation	Provide an interpretation of the key findings.			
16	Limitations	Discuss the limitations of the study, considering differences between the target trial and its emulation and the plausibility of assumptions, including assumptions regarding baseline confounding due to lack of randomization.			
Other information					
17	Ethics	Provide the institutional research board or ethics committee that approved the study and approval number(s), if relevant.			
18	Registration	State whether, when, and where the study protocol was registered.			
19	Sharing of study materials	Provide information on whether data, analytic code, and/or other materials are accessible and where and how they can be accessed.			
20	Funding source(s)	Provide the source(s) of funding and detail the role of the funder(s) in the design, conduct, and reporting of the study.			
21	Conflicts of interest	State any conflicts of interest and financial disclosures for all authors.			

[a]Republished with permission from the TARGET group. The TARGET Checklist is licensed by the TARGET group under the Creative Commons Attribution-No Derivatives (CC BY-ND) 4.0 International license.

Conclusions The TARGET guideline provides recommendations on reporting of studies explicitly emulating a target trial. Use of the guideline should facilitate transparency in reporting to improve peer review and help researchers, clinicians, and other readers interpret and use the results.

References

1. Hansford HJ, Cashin AG, Jones MD, et al. Development of the TrAnsparent ReportinG of observational studies Emulating a Target trial (TARGET) guideline. *BMJ Open*. 2023;13(9):e074626. doi:10.1136/bmjopen-2023-074626

2. 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

3. Hansford HJ, Cashin AG, Jones MD, et al. Reporting of observational studies explicitly aiming to emulate randomized trials. *JAMA Netw Open*. 2023;6(9):e2336023. doi:10.1001/jamanetworkopen.2023.36023

[1]School of Health Sciences, Faculty of Medicine & Health, University of New South Wales, Sydney, Australia, a.cashin.neura.edu.au; [2]Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, Australia; [3]CAUSALab, Harvard T.H. Chan School of Public Health, Boston, MA, US; [4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, US; [5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, US; [6]Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, US; [7]University of Exeter Medical School, Exeter, UK; [8]IQVIA, London, UK; [9]Richard A. and Susan F. Smith Center for Outcomes Research, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, US; [10]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; [11]Centre for Infectious Disease Epidemiology and Research, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa; [12]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; [13]RTI Health Solutions, Barcelona, Spain; [14]Northwestern University Feinberg School of Medicine, Chicago, IL, US; [15]Oxford Population Health, Big Data Institute, University of Oxford, Oxford, UK; [16]Faculty of Medicine, University of Southampton, Southampton, UK; [17]Department of Biostatistics, Boston University School of Public Health, Boston, MA, US; [18]Clinical Epidemiology & Biostatistics Unit, Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Melbourne, VIC, Australia; [19]Department of Paediatrics, The University of Melbourne, Parkville, Australia; [20]School of Population Health, Faculty of Medicine & Health, UNSW Sydney, Sydney, Australia; [21]Division of Pharmacoepidemiology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US; [22]Department of Public Health and Epidemiology, School of Population Health, RCSI University of Medicine and Health Sciences, Dublin, Ireland; [23]NIHR Bristol Biomedical Research Centre, Bristol, UK; [24]Health Data Research UK South-West, Bristol, UK; [25]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, US.

Conflict of Interest Disclosures Miguel A. Hernán is an advisor to ProPublica and Adigens Health, in which he owns equity, and a member of ADIA Lab's advisory board. Issa J. Dahabreh is a consultant to Moderna on work related to target trial emulation. Sebastian Schneeweiss is participating in investigator-initiated grants to Brigham and Women's Hospital from Boehringer Ingelheim, Takeda, and UCB unrelated to the topic of this study; is an advisor to and owns equity in Aetion, a software manufacturer; and is an advisor to Temedica, a patient-oriented data generation company.

Funding/Support This work received no direct funding. The TARGET consensus meeting was supported through proceeds from a course on causal inference (December 4-7, 2023; Sydney, Australia). Aidan G. Cashin was supported by an Australian National Health and Medical Research Council (NHMRC) Investigator Grant (2010088). Harrison J. Hansford was supported by an Australian NHMRC Postgraduate Scholarship (2021950), a PhD Top-Up Scholarship from Neuroscience Research Australia, and was a Neuroscience Research Australia PhD Pearl sponsored by Sandra Salteri, AO. Miguel A. Hernán was funded by a National Institutes of Health (NIH) grant (R37 AI102634). Issa J. Dahabreh was supported by a Patient-Centered Outcomes Research Institute (PCORI) award (ME-2021C2-22365) and an NIH award (R01HL136708). Barbra A. Dickerman was supported by a grant from the NIH (R00 CA248335). Matthias Egger was supported by grants from the NIH (R01 AI152772-01, 5U01-AI069924-05) and the Swiss National Science Foundation (32FP30-174281). Nazrul Islam was supported by grants from the National Institute for Health and Care Research (NIHR; HDRUK2022.0313) and the UK Office for National Statistics (2002563). Margarita Moreno-Betancur was supported by an Australian NHMRC Investigator Grant (2009572). Sebastian Schneeweiss was supported by the NIH (R01-HL141505, R01-AR080194) and the US Food and Drug Administration (HHSF223201710146C). Jonathan A. C. Sterne was supported by the NIHR Bristol Biomedical Research Centre and by Health Data Research UK. Elizabeth A. Stuart was supported by grants from the NIH (R01 MH126856) and PCORI (ME-2020C3-21145). James H. McAuley was supported by an Australian NHMRC Investigator Grant (2010128).

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 publication.

Additional Information Aidan G. Cashin and Harrison J. Hansford are co–first authors.

Acknowledgment We thank and acknowledge the contributions of all participants of the internal and external piloting who were not compensated for their contributions.

Author Practices and Experiences With PRISMA-P 2015

Mette B. Engmose,[1,2] An-Wen Chan,[3] Kerry Dwan,[4] Carsten Hinrichsen,[5] Asbjørn Hróbjartsson,[1,2] David Moher,[6] Matthew J. Page,[7] Larissa Shamseer,[8] Lesley A. Stewart,[4] Camilla H. Nejstgaard[1,2]

Objective Systematic review protocols often adhere incompletely to the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P) 2015.[1] Reporting guidelines need to be regularly updated to incorporate methodological developments, author feedback, and changed research context. We aimed to inform the update of PRISMA-P 2015[2] by exploring protocol authors' practices and experiences with the reporting guideline.

Design In a cross-sectional study, we investigated adherence to PRISMA-P 2015 and interviewed protocol authors. We randomly sampled 100 systematic review protocols from May 2021 to May 2024: 50 protocols published in PubMed-indexed journals and 50 protocols uploaded to or registered

in OSF/PROSPERO. Two authors independently assessed the 26 PRISMA-P 2015 items as fully, partially, or not reported (or not applicable). We analyzed protocol adherence overall and at the item level. We invited 43 corresponding authors from our sample of 100 systematic review protocols for interviews on their experiences with using PRISMA-P 2015. The selection of authors was based on maximum variation and snowball sampling and personal network. We used a piloted, semistructured interview guide involving 3 predefined themes: level of experience, views on using PRISMA-P 2015, and reflections on the guideline's strengths and weaknesses. We applied framework analysis to the interview transcripts. The study was reported according to Standards for Reporting Qualitative Research (SRQR).[3]

Results The PubMed-indexed protocols fully adhered to a median of 60% (range, 36%-80%) of PRISMA-P 2015 items. The equivalent median for OSF/PROSPERO protocols was 45% (range, 25%-88%). In both types of protocols, items not adhered to in more than 25% of the protocols were related to protocol amendments (item 4), role of the funder (item 5c), methods for planned summary other than quantitative (item 15d), meta-bias(es) (item 16), and confidence in the cumulative evidence (item 17) (**Table 25-1033**). In both types of protocols, 11 items partially adhered to in more than 25% of the protocols were lacking, eg, descriptions of procedures for data selection from multiple reports of the same study (item 11c). Of the 43 invited authors, 15 (9 men, 6 women) participated in the interviews. From the predefined themes, several suggestions for the PRISMA-P 2015 update emerged. Some suggestions regarded adding or modifying existing content, eg, to report conflicts of interest or how to report data synthesis when no meta-analysis was planned; other suggestions were more generic, eg, to add links to the Elaboration & Explanation paper.

Conclusions Adherence to PRISMA-P 2015 was inadequate in 50 systematic review protocols from PubMed-indexed journals and especially in 50 protocols from OSF/PROSPERO. Aspects often not adhered to were protocol amendments, role of the funder, methods for planned summary other than quantitative, meta-bias(es), and confidence in the cumulative evidence. The interviewed authors suggested several modifications of the guideline. Findings from this study will inform the update of PRISMA-P 2015.

References

1. Frost AD, Hróbjartsson A, Nejstgaard CH. Adherence to the PRISMA-P 2015 reporting guideline was inadequate in systematic review protocols. *J Clin Epidemiol*. 2022;150:179-187. doi:10.1016/j.jclinepi.2022.07.002

2. Nejstgaard CH, Shamseer L, Chan A-W, et al. Updating the PRISMA-P reporting guideline for systematic review protocols. OSF. April 10, 2024. https://osf.io/2znc5/

Table 25-1033. Adherence to PRISMA-P 2015 Items in Systematic Review Protocols Published in PubMed-Indexed Journals and Uploaded to or Registered in OSF/PROSPERO

Section and topic	Item	No. of protocols (%)							
		Fully adhered		Partially adhered		Did not adhere		Not applicable	
		PubMed	OSF/ PROSPERO	PubMed	OSF/ PROSPERO	PubMed	OSF/ PROSPERO	PubMed	OSF/ PROSPERO
Administrative information									
Identification	1a	50 (100)	34 (68)				16 (32)		
Update	1b		1 (2)					50 (100)	49 (98)
Registration	2	44 (88)		2 (4)		1 (2)		3 (6)	50 (100)
Contact	3a	50 (100)	44 (88)		4 (8)		2 (4)		
Contribution	3b	7 (14)	9 (18)	42 (84)	17 (34)	1 (2)	24 (48)		
Amendments	4	11 (22)	10 (20)	2 (4)	4 (8)	37 (74)	36 (72)		
Sources	5a	49 (98)	39 (78)			1 (2)	11 (22)		
Sponsor	5b	1 (2)	7 (14)	32 (64)	10 (20)	1 (2)	11 (22)	16 (32)	22 (44)
Role of the funder	5c	9 (18)	6 (12)			25 (50)	21 (42)	16 (32)	23 (46)
Introduction									
Rationale	6	49 (98)	41 (82)	1 (2)	5 (10)		4 (8)		
Objectives	7	42 (84)	40 (80)	8 (16)	9 (18)		1 (2)		
Methods									
Eligibility criteria	8	25 (50)	29 (58)	25 (50)	21 (42)				
Information sources	9	36 (72)	30 (60)	14 (28)	19 (38)		1 (2)		
Search strategy	10	43 (86)	35 (70)	7 (14)	7 (14)		8 (16)		
Data management	11a	5 (10)	3 (6)	38 (76)	41 (82)	7 (14)	6 (12)		
Selection process	11b	40 (80)	44 (88)	9 (18)	4 (8)	1 (2)	2 (4)		
Data collection process	11c	5 (10)	2 (4)	41 (82)	41 (82)	4 (8)	7 (14)		
Data items	12	33 (66)	33 (66)	17 (34)	13 (26)		4 (8)		
Outcomes and prioritization	13	32 (64)	20 (40)	18 (36)	25 (50)		5 (10)		
Risk of bias in individual studies	14	22 (44)	17 (34)	27 (54)	28 (56)	1 (2)	5 (10)		
Criteria for quantitative synthesis	15a	23 (46)	15 (30)	9 (18)	9 (18)	12 (24)	11 (22)	6 (12)	15 (30)
Methods of combining data	15b	33 (66)	12 (24)	8 (16)	18 (36)	2 (4)	5 (10)	7 (14)	15 (30)
Additional analyses	15c	39 (78)	20 (40)			5 (10)	15 (30)	6 (12)	15 (30)
Summary other than quantitative	15d	15 (30)	19 (38)	17 (34)	13 (26)	18 (36)	18 (36)		
Meta-bias(es)	16	15 (30)	9 (18)	21 (42)	18 (36)	14 (28)	23 (46)		
Confidence in the cumulative evidence	17	24 (48)	23 (46)		1 (2)	26 (52)	26 (52)		

3. O'Brien BC, Harris IB, Beckman TJ, et al. Standards for reporting qualitative research: a synthesis of recommendations. *Acid Med*. 2014;89(9):1245-1251. doi:10.1097/acm.0000000000000388

[1]Cochrane Denmark & Centre for Evidence-Based Medicine Odense (CEBMO), Department of Clinical Research, University of Southern Denmark, Odense, Denmark, mengmose@health.sdu.dk; [2]Open Patient data Exploratory Network (OPEN), Odense University Hospital, Odense, Denmark; [3]Department of Medicine, Women's College Research Institute, University of Toronto, Toronto, Ontario, Canada; [4]Centre for Reviews and Dissemination, University of York, York, UK; [5]National Institute of Public Health, University of Southern Denmark, Odense, Denmark; [6]Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada; [7]Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; [8]Knowledge Translation Program, Li Ka Shing Knowledge Institute, St Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada.

Conflict of Interest Disclosures An-Wen Chan, Kerry Dwan, Asbjørn Hróbjartsson, David Moher, Matthew J. Page, Larissa Shamseer, Lesley A. Stewart, and Camilla H. Nejstgaard are members of the steering group for the update of the PRISMA-P 2015 reporting guideline. All other authors declare no conflicts of interest related to financial interests, activities, relationships, and affiliations, including employment, affiliation, funding and grants received or pending, consultancies, honoraria or payment, speakers' bureaus, stock ownership or options, expert testimony, royalties, donation of medical equipment, or patents planned, pending, or issued. An-Wen Chan and David Moher are members of the Peer Review Congress Advisory Board but were not involved in the review or decision for this abstract.

Funding/Support This project did not receive any external funding; therefore, 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 were not influenced by any funder or sponsor.

Development of a Reporting Guideline on Health Equity in Observational Research (STROBE-Equity)

Cath Chamberlain,[1] Peter Craig,[2] Luis Gabriel Cuervo,[3] Omar Dewidar,[4] H. N. Ellingwood,[5] Elizabeth Ghogomu,[6] Billie-Jo Hardy,[7] Tanya Horsley,[8] Sonya Faber,[9] Cindy Feng,[10] Damian Francis,[11] Sarah Funnell,[12] Alison Krentel,[9] Janet Jull,[12] Elizabeth Kristjansson,[13] Julian Little,[9] Loveline Lum Niba,[14] Tamara Kredo,[15] Zack Marshall,[16] Lawrence Mbuagbaw,[17] Michael Johnson Mahande,[18] Stuart Nicholls,[19] Miriam Nkangu Nguilefem,[9] Ekwaro Obuku,[20] Oyekola Oloyede,[21] Ebenezer Owusu-Addo,[22] Kevin Pottie,[23] Jacqueline Ramke,[24] Alison Riddle,[6] Anita Rizvi,[13] Janet Hatcher Roberts,[9] Larissa Shamseer,[25] Melissa Sharp,[26] Janice Tufte,[27] Peter Tugwell,[28] Xiaoqin Wang,[29] Laura Weeks,[30] Charles Wisonge,[31] Luke Wolfenden,[32] Taryn Young,[33] Vivian Welch[6]

Objective Advancing health equity requires improvements in the conduct, analysis, and reporting of health research. Many observational studies are relevant to health equity because they use data that can advance our understanding of health inequities, including the systemic structures that perpetuate them. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline for the reporting of observational studies is highly endorsed by journal editors; however, the consideration of health equity has not been explicitly addressed in the STROBE statement. As a result, editors have been limited in their ability to promote the reporting of health equity data and considerations. The STROBE-Equity extension provides actionable guidance for equity-relevant observational studies to fill this gap.

Design The development of STROBE-Equity adhered to a published protocol, informed by the Enhancing the Quality and Transparency of Health Research (EQUATOR) Network's methodological framework including a scoping review of existing guidance, methods study, key informant interviews, and a global online survey. The project team was diverse across multiple domains, including lived experience of health inequity, gender, career stage, age, ethnicity, Indigenous background, disciplines, place of residence, and different decision-making roles. We recruited team members from our international network of patients, researchers, editors, and practitioners who had prior involvement in observational studies or expertise in reporting guidelines, biostatistics, knowledge translation, and patient engagement. We applied an integrated knowledge translation approach to (1) assess the reporting of health equity in published observational studies through a scoping review of available guidance, methodological review of observational studies, and key informant interviews; (2) seek wide international input on draft items; and (3) establish consensus amongst knowledge users and researchers through a 2-day consensus meeting. The development of the reporting guideline occurred between January 2020 and February 2024, with consensus achieved in May 2024.

Results Our process identified 10 extension items for the STROBE statement to offer guidance for reporting of health equity (**Table 25-1136**). The items involve the following: (1) facilitating the identification of health equity relevance, (2) describing the purpose of the study in relation to health equity, (3) describing team composition, (4) engagement with individuals experiencing inequities relevant to the condition of interest, (5) sampling and recruitment of relevant populations, (6) participant selection, (7) analytical methods, (8) describing participant flow for relevant populations, (9) describing the characteristics of study participants, and (10) interpretation of study findings in relation to health equity.

Conclusions The use of the STROBE-Equity extension alongside the original STROBE statement aims to promote transparent reporting of health equity data and considerations in observational research. Journal editors endorsing the use of this guideline can help advance health equity through improved reporting, ultimately contributing to research that informs more equitable policies and interventions.

[1]University of Melbourne, Australia; [2]MRC/CSO Social and Public Health Sciences Unit, School of Health and Wellbeing, University

Table 25-1136. STROBE-Equity Extension Items to Enhance Reporting of Health Equity in Observational Studies

STROBE section	STROBE-Equity extension items
Title and abstract	1. Highlight the relevance of the study to health equity
Methods	2. Describe the study rationale as it pertains to health equity 3. Report research team composition 4. Describe engagement with individuals with lived experience 5. Describe sampling and recruitment methods 6. Describe study participant selection methods
Results	7. Describe health equity analytical methods 8. Report the flow of participants according to characteristics associated with inequities 9. Report characteristics of study participants according to characteristics associated with inequities
Discussion	10. Interpret results while considering study sample representativeness and implications for health equity

of Glasgow, UK; [3]Pan American Health Organization, Washington, DC, US; [4]University of Toronto, Canada; [5]Carleton University, Canada; [6]Bruyère Health Research Institute, Canada; [7]Dalla Lana School of Public Health, University of Toronto, Canada; [8]Royal College of Physicians and Surgeons, Canada; [9] School of Epidemiology and Public Health, University of Ottawa ,Canada; [10]Dalhousie University, Canada; [11]PhD/School of Health and Human Performance, Center for Health and Social Issues, Georgia College, Milledgeville, GA, US; [12]Queens University, Canada; [13]School of Psychology, University of Ottawa, Canada; [14]Department of Public Health, Faculty of Health Sciences, The University of Bamenda, Cameroon; [15]South Africa Medical Research Council, South Africa; [16]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary and O'Brien Institute for Public Health, University of Calgary, Canada; [17]McMaster University, Canada; [18]Kilimanjaro Christian Medical Centre, Kenya; [19]Centre for Practice-Changing Research, Ottawa Hospital Research Institute, Canada; [20]Africa Centre for Systematic Reviews and Knowledge Translation, Kampala, Uganda; [21]Sefako Makgatho Health Sciences University, South Africa; [22]Bureau of Integrated Rural Development, College of Agriculture and Natural Resources, Ghana; [23]Dalhousie University, Canada; [24]International Centre for Eye Health, London School of Hygiene & Tropical Medicine, London, UK; [25]Knowledge Translation Program, Li Ka Shing Knowledge Institute, Unity Health Toronto, Canada; [26]Department of Public Health & Epidemiology, School of Population Health, Ireland; [27]Catalyst for Positive Change, US; [28]Department of Medicine, Faculty of Medicine, University of Ottawa, Canada; [29]University of Ottawa Heart Institute, Canada; [30]Canada's Drug Agency, Ottawa, Canada; [31]South Africa Medical Research Council, South Africa; [32]University of Newcastle, Australia; [33]Stellenbosch University, South Africa; vwelch@campbellcollaboration.org.

Conflict of Interest Disclosures Kevin Pottie, Lawrence Mbuagbaw, Vivian Welch, and Peter Tugwell declare they are co-convenors of the Cochrane Health Equity Thematic Group. No other disclosures were reported.

Funding This work was funded by the Canadian Institutes of Health Research. The funder had no role in the project design, implementation, or decision to publish. Vivian Welch is funded by a Public Health Agency of Canada and Canadian Institute of Health Research Applied Public Health Chair award. Kevin Pottie is funded by a Dalhousie University Faculty of Medicine Chair in Family Medicine.

Acknowledgment The authors thank all the respondents to the global survey.

Reproducibility

In-person

Prevalence of Reproducible Health Sciences Research: A Systematic Review and Meta-Analysis

Niklas Bobrovitz,[1,2] Harriet Ware,[2] Corson Johnstone,[3] Juliane Kennett,[3] Stephana Julia Moss,[4] Liam Whalen-Browne,[3] Faizan Khan,[5] Benjamin Fletcher,[2] Daniel J. Niven,[3,6,7] Henry T. Stelfox[3,6,8]

Objective To synthesize evidence on the prevalence and measurement of reproducibility in health sciences research.

Design This was a systematic review with meta-analysis registered a priori.[1] We searched MEDLINE, Embase, Web of Science, the Cochrane Controlled Trials Registry and Database of Systematic Reviews, and grey literature published from January 1, 2000, to May 31, 2023. We included English-language articles in the health sciences that reported the prevalence of reproducible research or the prevalence of studies dedicated to reproducing research. We included studies published after 2000 to ensure relevance to current practice. Two authors independently screened articles in duplicate. Extractions were completed by one reviewer and verified by a second. The prevalence of reproducible research was defined as the proportion of research findings or studies for which the methods could be exactly repeated (methodological), the results could be corroborated by a study using the same or similar methods (results), or qualitatively similar conclusions could be made based on a study reproduction (inferential). Risk of bias was assessed using a modified Joanna Briggs Institute checklist. We meta-analyzed estimates using random effects models. To explore between-study heterogeneity, we conducted univariable and multivariable meta-regression using generalized linear mixed models. The level of significance was set at $P < .05$.

Results A total of 177 studies published between 2001 and 2023 were included, covering all health sciences domains: biomedical/preclinical (n = 56), clinical (n = 106), health/social care services (n = 2), population health (n = 5), and multiple domains (n = 8). Studies reported on 3 types of reproducibility: methodological (n = 68), results (n = 129), and inferential (n = 5). There were 7 methodological approaches for conducting reproducibility research and 38 metrics to quantify reproducibility. The pooled estimated prevalence of reproducible research was 36.2% (95% CI, 29.7%-43.3%; I^2 = 99.8%; n = 180 estimates) with variation by research domain: 30.3% (95% CI, 20.3%-42.6%; n = 59 estimates) for biomedical/preclinical, 38.3% (95% CI, 30.0%-47.4%; n = 106 estimates) for clinical, 66.6% (95% CI, 20.7%-93.9%; n = 3 estimates) for health/social care services, and 36.5% (95% CI, 0.0%-100.0%; n = 3 estimates) for population health. There was a nonsignificant decline over time between 2001 and 2023 of −4.7% (95% CI −10.1 to 1.1%; P = .11) per year in the odds of reproducibility (**Figure 25-1041**). The pooled estimated prevalence of studies

dedicated to reproducing research was 1.4% (95% CI, 0.40%-4.10%; I^2 = 99.2%; n = 40 estimates). Most estimates were at moderate risk of bias (n = 193), with some at high (n = 25) or low (n = 2) risk. Meta-regression of publication year, research domain, type of reproducibility, metric to quantify reproducibility, and risk of bias explained minimal estimate heterogeneity.

Conclusions We found that an estimated one-third of health sciences research was reproducible and 1 in 100 studies were dedicated to reproducing research. Further investigation is needed to understand heterogeneity in estimates. Standardized strategies for quantifying reproducibility should be developed. Reproducibility studies are needed in health/social care services and population health research.

Reference

1. Bobrovitz N. The reproducibility of health research. Open Science Framework. 2021. https://osf.io/3fhd9

[1]Department of Emergency Medicine, Cumming School of Medicine, University of Calgary, Canada, njhbobro@ucalgary.ca; [2]Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Canada; [3]Department of Critical Care Medicine, Cumming School of Medicine, University of Calgary, Canada; [4]Department of Pediatrics, Division of Infectious Diseases, IWK Health Centre, Dalhousie University, Canada; [5]Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Canada; [6]O'Brien Institute for Public Health, University of Calgary, Canada; [7]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Canada; [8]Faculty of Medicine and Dentistry, University of Alberta, Canada.

Conflict of Interest Disclosures None reported.

Funding/Support The work was funded by the Canadian Institutes of Health Research.

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

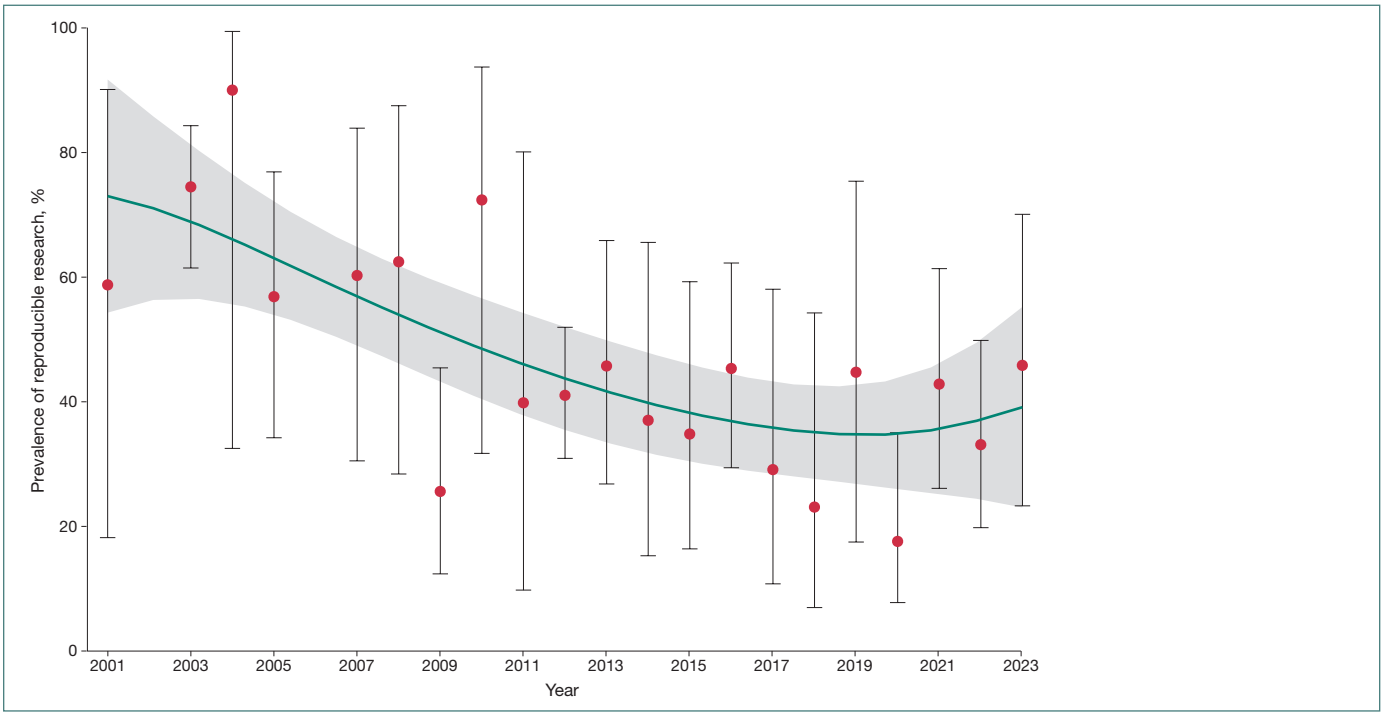
Factors Associated With the Reproducibility of Health Sciences Research: A Systematic Review and Evidence Gap Map

Stephana Julia Moss,[1] Juliane Kennett,[2] Jeanna Parsons Leigh,[3] Niklas Bobrovitz,[4] Henry T. Stelfox[5]

Objective To map the evidence for factors (eg, research practices) associated with the reproducibility of methods and results reported in health sciences research.

Design Five bibliographic databases were searched from January 2000 to May 2023, followed by supplemental searches of high-impact journals and relevant records. We included health science records of observational, interventional, or knowledge synthesis studies reporting data on factors related to research reproducibility. Factors were operationalized as modifiable or nonmodifiable aspects of study conduct relating to individual-, study-, or institutional-level practices, methods, and processes that could impact the reproducibility of research methods or results.[1] Reproducibility was operationalized by 2 mutually exclusive categories: (1) methodological reproducibility (ie, the ability to exactly repeat the methods, including study procedures and data analysis) and (2) results reproducibility (ie, obtaining corroborating results using the same or similar methods).[2] We included studies that used surrogate measures for reproducibility (eg, type 1 or 2 error rates) if they (1) explicitly stated their aim to investigate the reproducibility of research and (2) rationalized their choice of surrogate measure.[3] Data were coded using inductive qualitative

Figure 25-1041. Annual Trend in Meta-Analyzed Estimates of Prevalence of Reproducible Health Sciences Research



Bars indicate 95% CIs, dots indicate point estimates, and grey shading indicates 95% smoothing CI. The green line indicates polynomial smoothing.

content analysis, and empirical evidence was synthesized with evidence and gap maps. Study risk of bias was assessed using the Quality in Prognostic Studies risk-of-bias tool. Statistical tests of the association between factors and reproducibility outcomes were summarized as reported in the included articles.

Results Our review included 148 primarily biomedical and preclinical (n = 62) and clinical (n = 71) studies. Factors were classified into 12 modifiable (eg, sample size and power) and 3 nonmodifiable (eg, publication year) categories. Of 234 reported evaluations of factors, 76 (32%) assessed methodological reproducibility and 158 (68%) assessed results reproducibility. The most frequently reported factor was transparency and reporting (38 of 234 assessments [16%]). A total of 155 factors (66%) were evaluated for statistical associations with reproducibility outcomes (**Table 25-0858**). Statistical associations were most frequently conducted for analytical methods (24 of 26 reporting significance [92%]), sample size and power (21 of 23 reporting significance [91%]), and participant characteristics and study materials (10 of 12 reporting significance [83%]). Risk-of-bias assessments found low risk of bias for study participation, factor measurement, and statistical analysis, and high risk of bias for confounding.

Conclusions Our review identified a large body of literature consisting primarily of observational studies of factors associated with the reproducibility of health sciences research. The data suggest that reproducibility may be improved by implementing more stringent statistical testing procedures and thresholds, sample size and power calculations, and improved transparency and completeness of reporting. Experimental studies are needed to test interventions to improve reproducibility. Factors identified in this study with consistent observational support should be prioritized for experimentation. Factors that affect reproducibility in health and social care services and population and public health need to be identified given the paucity of data in these areas.

Table 25-0858. Statistically Significant Associations Between Factors and Reproducibility Outcomes

Factor category	Reproducibility outcomes, No. (%)		
	Total frequency, N = 234 (100%)	Methods, n = 76 (32%)	Results, n = 158 (68%)
Modifiable factors	**n = 114 (49%)**	**n = 28 (37%)**	**n = 86 (54%)**
Transparency and reporting, n = 38 (16%)	9 (64)	5 (71)	4 (57)
Analytical methods, n = 33 (14%)	24 (92)	3 (100)	21 (91)
Study rigor and quality, n = 27 (11%)	12 (44)	1 (50)	11 (73)
Sample size and power, n = 25 (11%)	21 (91)	6 (100)	15 (88)
Data and code sharing, n = 18 (7%)	1 (33)	0	1 (33)
Participant characteristics and study materials, n = 14 (6%)	10 (83)	2 (100)	8 (80)
Physical research environment, n = 12 (5%)	6 (67)	2 (67)	4 (67)
Culture and incentives, n = 6 (3%)	0	0	0
Software platforms, n = 6 (3%)	1 (50)	0	1 (50)
Specialist involvement, n = 5 (2%)	3 (75)	3 (75)	0
Training and supervision, n = 4 (2%)	0	0	0
Publication bias, n = 2 (1%)	2 (100)	0	2 (100)
Nonmodifiable factors	**n = 41 (18%)**	**n = 11 (14%)**	**n = 30 (19%)**
Academic success metrics, n = 20 (9%)	11 (58)	2 (50)	9 (60)
Research discipline, n = 14 (6%)	8 (57)	3 (75)	5 (50)
Publication year, n = 10 (4%)	4 (50)	2 (67)	2 (40)

Sorted by frequency of total factors. Individual studies could have assessed more than 1 factor category. Values represent counts and percentages of studies that assessed for associations and found a statistically significant association.

Conflict of Interest Disclosures None reported.

Funding/Support This work was funded by the Canadian Institutes of Health Research.

Role of the Funder/Sponsor The funding body 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.

References

1. Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? *Sci Transl Med*. 2016;8(341):341ps12-341ps12. doi:10.1126/scitranslmed.aaf5027

2. Niven DJ, McCormick TJ, Straus SE, et al. Reproducibility of clinical research in critical care: a scoping review. *BMC Med*. 2018;16:1-12. doi:10.1186/s12916-018-1018-6

3. Clemens MA. The meaning of failed replications: a review and proposal. *J Econ Surveys*. 2017;31(1):326-342. doi:10.1111/joes.12139

[1]Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada, sj.moss@dal.ca; [2]Department of Critical Care Medicine, University of Calgary, Calgary, Alberta, Canada; [3]Faculty of Health, Dalhousie University, Halifax, Nova Scotia, Canada; [4]Department of Emergency Medicine, University of Calgary, Calgary, Alberta, Canada; [5]Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada.

Perception of Open Science Practices on Reproducibility Among Reviewers of Grant Proposals at Research Funding Organizations

Ayu Putu Madri Dewi,[1] Nicholas J. DeVito,[2] Gowri Gopalakrishna,[1,3] Inge Stegeman,[4,5] Mariska Leeflang[1]

Objective To improve research reproducibility, referees of grant proposals could assess the expected reproducibility of a proposed project. A simple Open Science checklist may assist referees in assessing whether referees are more capable of predicting the reproducibility of research proposals with an Open Science checklist than without one.[1]

Design This is a nonrandomized controlled trial and a comparative prediction study using a minimum of 100 granted research proposals and their corresponding published outputs (α = 5%, power = 80%, correlation = 0.6).[2] The study focuses on empirical research with quantitative data across diverse fields. A mock grant review was started on April 7, 2025, with a minimum of 4 referees divided into 2 groups: one using an Open Science checklist and the other not. Each proposal will be assessed under both conditions (**Figure 25-0934**). The checklist, developed in a previous study[3] and adapted for grant proposal evaluation, serves as the intervention. Referees are blinded to final study outcomes to minimize bias. The primary outcome is predictive accuracy, assessed using the area under the curve (AUC) from receiver operating characteristic (ROC) analysis. Referees' predictions on the actual reproducibility of the published studies will be compared. Open Science scores are numerical (0-14), while reproducibility outcomes are binary (yes or no). The AUC from logistic regression will be used and visualized in ROC curves. Comparative predictive accuracy will be calculated for both groups. Sensitivity analyses will explore variations in checklist scoring and study reproducibility across different research fields. A database including grant proposal data, Open Science scores, reproducibility status, and analytical code in R will be shared in an open repository. We hypothesize that referees without the checklist will predict reproducibility no better than random chance (with an AUC of approximately 0.50), while those using the checklist will achieve an AUC of approximately 0.65, indicating improvement while acknowledging that the checklist is not a perfect tool.

Results As of June 5, 2025, a total of 15 mock grant reviewers have completed 89 mock reviews across 60 unique grant proposals: 46 reviews in the checklist group and 43 in the control group. Twenty-nine proposals were reviewed in both groups, forming a fully paired subset. Analysis is ongoing.

Conclusions We will share our results with funding agencies to explore integrating the Open Science checklist into grant review processes to enhance reproducibility.

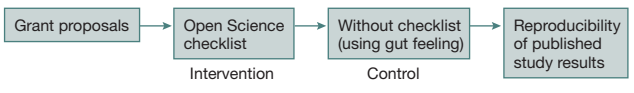
References

1. Parker TH, Griffith SC, Bronstein JL, et al. Empowering peer reviewers with a checklist to improve transparency. *Nat Ecol Evol*. 2018;2(6):929-935. doi:10.1038/s41559-018-0545-z

2. Fraser H, Bush M, Wintle BC, et al. Predicting reliability through structured expert elicitation with the repliCATS (Collaborative Assessments for Trustworthy Science) process. *PLoS One*. 2023;18(1):e0274429. doi:10.1371/journal.pone.0274429

3. Dewi APM, Rethlefsen M, Schroter S, et al. Measuring the efficacy of an intervention to improve reproducibility of scientific manuscripts: study protocol for a randomized controlled trial. OSFHOME. October 31, 2024. Accessed July 2, 2025. https://osf.io/b5g6y

Figure 25-0934. Study Flow



Please note that conducting the intervention first before the control is not necessary. All proposals will undergo the same assessment in both the intervention and control groups.

[1]Amsterdam UMC, Epidemiology and Data Science Department, Amsterdam, The Netherlands; [2]Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK, nicholas.devito@phc.ox.ac.uk; [3]Department of Epidemiology, Faculty of Health, Medicine, and Life Sciences Maastricht University, The Netherlands; [4]Department of Otorhinolaryngology and Head & Neck Surgery University Medical Center Utrecht, The Netherlands; [5]Brain Center, University Medical Center Utrecht, The Netherlands.

Conflict of Interest Disclosures None reported.

Virtual

Testing Computational Reproducibility Review in Editorial Workflows of Academic Journals: A Randomized Controlled Trial From the European iRISE Project

Laura Caquelin,[1] Rachel Heyard,[2] Stephanie Zellers,[3] Hanno Würbel,[4] Gustav Nilsonne[1]

Objective Computational reproducibility is vital to research quality, enabling others to replicate analyses and achieve the same results. Despite its importance, it is rarely assessed systematically. This study aims to evaluate whether computational reproducibility review during the publication process improves reproducibility compared with standard peer review. We hypothesize that this intervention will improve reproducibility, encourage code sharing, and reduce errors.

Design This randomized controlled trial was designed to enroll manuscripts submitted to partnering journals that meet inclusion criteria, including open data availability and inferential analysis. Manuscripts will be randomized 1:1 to peer review with computational reproducibility review (intervention group) or standard peer review without intervention (control group). The computational reproducibility review involves reproducing the essential statistical results using shared data (with or without the code). Feedback is provided to authors in the intervention group during the peer review process. The reproducibility of the manuscripts randomized to the control group will be assessed after publication only. The primary outcome is the proportion of manuscripts for which we successfully reproduced the essential statistical results after publication. Secondary outcomes include rates of code sharing, overt errors identified, time required for the review, publication timelines, and categorized reproducibility issues.

Results This study began in January 2025, with a preregistered protocol openly available on the Open Science Framework.[1] Recruitment is ongoing, and as of April 2025, 28

manuscripts submitted to the partnering journal GigaSicence have been screened, with 17 deemed eligible. The project is scheduled to run through September 2026, with the goal of including 200 manuscripts. This presentation will outline key insights gained during the study's implementation and preliminary results.

Conclusions This study fills an important gap in assessing research quality by testing a practical way to improve computational reproducibility. If successful, the results could support wider use of reproducibility reviews in academic journals, leading to more reliable and trustworthy science.

Reference

1. Caquelin L, Heyard R, Zellers S, et al. A study protocol for an intervention study using computational reproducibility testing to improve reproducibility. Open Science Framework. Updated January 29, 2025. doi:10.17605/OSF.IO/3Y8FP

[1]Karolinska Institutet, Stockholm, Sweden, laura.caquelin@ki.se; [2]Center for Reproducible Science at the Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland; [3]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; [4]Division of Animal Welfare, University of Bern, Bern, Switzerland.

Conflict of Interest Disclosures None reported.

Funding/Support This study is part of the iRISE (Improving Reproducibility in Science) project, which has received funding from the European Union under the Horizon Europe program (grant agreement 101094853).

Role of the Funder/Sponsor The funder had no role in the data collection, analysis, or conclusions of the abstract.

Research Methods

In-person

Characterizing Adverse Event Methods Reported in ClinicalTrials.gov and Publications

Kyungwan Hong,[1] Mark Basista,[1] Tony Tse[1]

Objective Information about how adverse events (AEs) were specified, collected, and assessed for clinical trials (ie, AE methods) is essential to accurately interpret the safety profiles of studied interventions.[1] ClinicalTrials.gov provides a structured format for trialists to report aggregate results information and AE methods that complement the published literature. Prior work compared reporting of AEs in publications and ClinicalTrials.gov in terms of completeness and consistency,[2] examined differences between sources in reporting AE methods for systematic collection,[3] and assessed AE methods reporting in publications and identified methodologic weaknesses.[1] This cross-sectional analysis explored the extent that AE methods reported in ClinicalTrials.gov complemented corresponding publications.

Design ClinicalTrials.gov was searched on August 1, 2024, for drug and biologic trials with primary completion dates between January 18, 2017, and August 1, 2021, to allow 3 years for results posting and publication. A total of 9659 records were initially retrieved. After randomization, the records were reviewed sequentially until a convenience sample of 40 trials with summary results posted on ClinicalTrials.gov and corresponding primary results articles in PubMed were identified. The identified trials were then analyzed. All authors reviewed and extracted text for 5 ClinicalTrials.gov AE method data elements: collection period (time frame), vocabulary source for AE terms (how AE terms were standardized), collection approach (systematic or nonsystematic), analysis population, and whether participants withdrew due to harms. Each set of trial results was independently reviewed by 2 authors, with discrepancies resolved by consensus.

Results Collection period (28 [70.0%]), analysis population (27 [67.5%]), and withdrawal for harm (35 [87.5%]) were described for over half of the 40 sampled trials in ClinicalTrials.gov and/or publications (**Table 24-0831**). However, consistency varied; for example, of the 35 trials with number of participants withdrawn for harm described in both sources, 20 (57.1%) provided inconsistent information. Collection approach was described in both sources infrequently (4 [10.0%]). Other AE methods (eg, vocabulary source) were reported for some trials in only one or neither source.

Conclusions A few AE methods were not reported on ClinicalTrials.gov and/or publications, indicating a need for improvement. Complete AE methods reporting is needed for accurate safety profiles of study interventions and unbiased risk-benefit assessments. Even when available, reported AE methods were frequently inconsistent between sources, potentially due to variations in time frames used for AE reporting. Limitations included excluding supplementary appendices and study protocols that may provide additional methodologic details on AE methods. Future research is needed to review publicly available study protocols for AE methods information, compare the quality of AE methods reporting by source, and characterize trials in which AE methods have and have not been completely reported (eg, by funding source, study design).

References

1. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ Open*. 2019;9(2):e024537. doi:10.1136/bmjopen-2018-024537

2. Krešo A, Grahovac M, Znaor L, Marušić A. Safety reporting in trials on glaucoma interventions registered in ClinicalTrials.gov and corresponding publications. *Sci Rep*. 2024;14(1):27762. doi:10.1038/s41598-024-79394-z

3. Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K; MUDS Investigators. 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 24-0831. Availability and Consistency of AE Methods Data Elements in Drug or Biologic Trials Registered in ClinicalTrials.gov and Matching Publication[a]

AE method element	Availability by source, No. (%)						Overall availability, No. (%)	
	ClinicalTrials.gov only	Publication only	Both			Neither	ClinicalTrials.gov	Publications
			Total	Consistent	Inconsistent			
Collection period[b]	12 (30.0)	0	28 (70.0)	14 (50.0)	14 (50.0)	0	40 (100)	28 (70.0)
Collection approach[b]	36 (90.0)[c]	0	4 (10.0)	2 (50.0)	2 (50.0)	0	40 (100)[c]	4 (10.0)
Vocabulary source	14 (35.0)	1 (2.5)	20 (50.0)	10 (50.0)	10 (50.0)	5 (12.5)	34 (85.0)	21 (52.5)
Analysis population description[d]	1 (2.5)	4 (10.0)	27 (67.5)	21 (77.8)	6 (22.2)	8 (20.0)	28 (70.0)	31 (77.5)
Withdrawal for harm[d]	2 (5.0)	3 (7.5)	35 (87.5)	15 (42.9)	20 (57.1)	0	37 (92.5)	38 (95.0)

Abbreviation: AE, adverse event.

[a]Of the 40 sampled trials, most were industry funded (33 [82.5%]) and studied US-regulated drugs or biologics (33 [82.5%]); median study participant size was 251 (IQR, 68-524). Of 35 trials with multiple arms, 32 (91.4%) were randomized and 24 (68.6%) involved blinding.

[b]Required information on ClinicalTrials.gov.

[c]Three trial records that did not list any observed AEs also did not report the AE collection approach.

[d]Extraction of information from ClinicalTrials.gov was not limited to the AE information module but potentially included information reported in the ClinicalTrials.gov participant flow and outcome measures modules.

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, US, kyungwan.hong@nih.gov.

Conflict of interest Disclosures Kyungwan Hong reported being employed by the National Institutes of Health (NIH) as a policy analyst with the ClinicalTrials.gov program. Mark Basista reported conducting this work while contracting with ICF. Tony Tse reported being employed by the NIH as an analyst with the ClinicalTrials.gov program. No other disclosures were reported.

Funding/Support This work was supported by the National Center for Biotechnology Information of the National Library of Medicine, NIH.

Role of the Funder/Sponsor The funder was not directly involved 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 views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the NIH.

Bias in Machine Learning Associated With Weak Baselines, Data Leakage, and Inadequate Measures Reporting

Randall J. Ellis,[1] Chirag J. Patel[1]

Objective New omics modalities hold promise for biomarker discovery for disease prediction. Accessible biobank samples have fueled a massive literature, some of which may be irreproducible, be subject to lack of generalizability, or report inflated results due to data leakage,[1] ie, spurious relationships between input and target variables that arise as artifacts of data collection, sampling, or preprocessing. These biases frequently result in models developed in one context not generalizing to real-world contexts. Here, we demonstrated and quantified bias in biomedical multiomic machine learning results due to irrelevant or unreported baselines, data leakage, and inadequate performance metrics reporting.

Design In November 2024, using data from the UK Biobank, we conducted a cohort study of all 607 disease outcomes defined in the dataset, using blood plasma proteomics data (collected from 2006-2010) and demographics data. We demonstrated how simple demographics (age, sex, education) compare competitively with putatively novel models that incorporate an expansive number of omics input features, how randomly chosen omics features compare with those chosen according to data-driven feature selection methods, the impact of data leakage on performance (eg, normalizing the data before making train-test cross-validation splits), and how presenting the area under the receiver operating characteristic curve (AUROC) gives a biased view of performance. We presented additional case studies of using baseline risk calculators for cardiovascular disease and atherosclerotic cardiovascular disease to assess how omics factors compared with traditional risk scores for associations with 10-year risk of heart disease or stroke. We used a stratified cross-validation approach and assessed the AUROC, sensitivity, specificity, and positive and negative predictive value.

Results Across a majority of 607 disease outcomes, demographic baselines performed competitively (0%-10% difference in mean AUROC) in comparison with the combination of demographics and omics factors when looking at the variability of performance across cross-validation folds, which are underreported in the literature. We found that data leakage influenced predictive performance (mean AUROC increases of 5%-30%). Accounting for disease prevalence and class balance played a significant role in the interpretation of machine learning results, particularly for positive predictive value. Omics factors contributed marginal increases in performance for cardiovascular disease and atherosclerotic cardiovascular disease compared with traditional risk score calculators (less than 5% mean AUROC).

Conclusions The robustness and transparency of biobank-based omics research is enhanced if several practices are followed: (1) rigorous and relevant baselines would provide critical tests of candidate models to justify their clinical use and value, (2) preventing data leakage would preclude inflated results, and (3) comprehensive performance metrics

reporting would give transparent measures of progress for researchers in the omics community. Our results complement other work showing the impact of data leakage[1] and the marginal improvements of plasma proteomics beyond demographic baselines.[2]

References

1. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*. 2023;4(9):100804. doi:10.1016/j.patter.2023.100804

2. Deng YT, You J, He Y, et al. Atlas of the plasma proteome in health and disease in 53,026 adults. *Cell*. 2025;188(1):253-271. doi:10.1016/j.cell.2024.10.045

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, US, chirag_patel@hms.harvard.edu.

Conflict of Interest Disclosures None reported.

Funding/Support This work was supported by grants from the National Institute on Aging (RF1AG074372) and National Library of Medicine (5T15LM007092-33).

Role of the Funder/Sponsor The funders were not involved in the development or implementation of this study.

Acknowledgment We acknowledge the support of our funders and Audrey Airaud for early-stage input on these experiments.

Individual-Participant Data Meta-Analysis Methodological Guidance: A Systematic Review

Edith Otalike,[1] Mike Clarke,[2] Ngianga-Bakwin Kandala,[1] Joel J. Gagnier[1,3]

Objective Individual-participant data meta-analysis (IPD-MA) is regarded as the criterion standard in evidence synthesis, but it is resource intensive. While there is consensus on reporting items, a consensus-based tool for the critical appraisal of the methodological quality of IPD-MA does not currently exist. We undertook a systematic methodology review as the initial phase in the development of a critical appraisal checklist. This review collated and summarized the available methodological guidance on IPD-MA of randomized and observational studies.

Design We followed the guidelines for Cochrane Methodology Reviews and reported following the PRISMA 2020 guidance. We performed an electronic search of MEDLINE, Embase, CINAHL, Web of Science Scopus, Cochrane Methodology Registry, CONSORT Database of Methodological Papers, Health Technology Assessment Review Database, *Research Synthesis Methods*, and *Journal of the Royal Statistical Society* covering publications from 1946 to June 2024. We included studies published in English that addressed any methodological guidance and essential statistical and software requirements for IPD-MA. Data extraction focused on study characteristics, domain of the review process, and the specific recommendations. Risk of bias was assessed using resources relevant to the study design. A thematic synthesis was performed to group recurring themes into domains. For each domain, signalling questions were generated to develop a preliminary checklist for assessment and refinement in a 2-round e-Delphi survey involving international IPD-MA experts.

Results The literature search yielded 13,589 citations. After screening 9436 unique abstracts and reviewing 286 full texts, we included 130 articles that met our inclusion criteria. These articles consisted of narrative reviews, handbooks, critical reviews, empirical studies, and statistical method articles. They were published between 1995 and 2024, with most originating from the UK (62 [48%]), the US (20 [15%]), and the Netherlands (16 [12%]) and 50 (38%) originating from 10 other countries. Most of these studies had a low risk of bias. We identified 14 domains of guidance for conducting and reporting of IPD-MA, and we categorized them into 5 sections (**Table 25-0999**). This finding informed the initial version of the checklist to be evaluated in the e-Delphi survey.

Conclusions There are many recommendations in the literature on the general conduct of IPD-MA and on specific aspects of this research, which would benefit from consensus recommendations for all aspects of IPD-MA and critical appraisal of reports. This review provides many suggestions for these recommendations, and our e-Delphi survey will seek consensus on items to include in the critical appraisal tool, which we expect to be completed in 2025.

[1]Department of Epidemiology & Biostatistics, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada, eotalike@uwo.ca; [2]Northern Ireland Methodology Hub, Queen's University Belfast, Northern Ireland, UK; [3]Department of Surgery, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada.

Conflict of Interest Disclosures Edith Otalike receives internal funding from Western University and the Dean's Research Scholarship award. No other disclosures were reported.

Additional Information Joel J. Gagnier is a co–corresponding author (jgagnie4@uwo.ca).

Reporting of Confounder Selection in Observational Studies in High-Impact Medical and Epidemiological Journals, 2003-2023

Luis C. L. Correia,[1] Rafael F. Mascarenhas,[2] Felipe S. C. de Menezes,[2] Jeronimo S. Oliveira Júnior,[2] Marcus V. Almeida,[2] Caio F. Azevedo,[2] Naieli M. de Andrade,[2] Viola Vaccarino,[1] Joseph S. Ross,[3] Joshua D. Wallach[1]

Objective Several approaches exist for selecting confounders for adjustment in observational studies, including subjective judgment, statistical criteria, or the use of a causal model, usually represented by directed acyclic graphs (DAGs). We assessed the characteristics and trends in the reported methods for selecting confounders to control for in observational studies published in the highest impact factor medical and epidemiological journals.

Design We identified the 10 highest Impact Factor medical (n = 5) and epidemiological (n = 5) journals according to InCites Journal Citation Reports. For each journal, we reviewed all PubMed-indexed articles published in 2003,

Table 25-0999. List of Identified Domains With Possible Signalling Questions

Domain
Section 1. Systematic review framework
1. Availability of review protocol
Was a review protocol established before the IPD-MA was conducted and made publicly available?
2. Identification of eligible studies
Was there a comprehensive process for identifying eligible studies?
3. Availability of IPD
What proportion of the participants in the eligible studies was IPD retrieved for?
Section 2: Data collection and processing
4. Data processing
Were outcome definitions and any standardizations uniform across all intervention groups in all studies?
5. Data integrity and quality assessment
Were key aspects of the IPD checked?
6. Risk of bias assessment
Was an appropriate risk of bias tool used to assess study quality based on reports of the included studies (rather than the IPD)?
Section 3: Statistical analysis
7. Combining IPD and aggregate data
Was an appropriate method used to combine IPD and aggregate data results in the MA?
8. MA
Did the reviewers stratify the analysis to account for the clustering of study participants?
9. Missing data
Were missing data appropriately handled in the analysis?
10. Subgroup analysis
Was an appropriate method used to estimate effect variation at the participant level?
11. Heterogeneity
Was clinical heterogeneity considered, even in the absence of statistical heterogeneity?
Section 4: Interpretation and reporting
12. Interpretation of findings
Did the interpretation consider the analytic approach taken for the IPD-MA?
13. Reporting
Did the review report follow the guidance in the PRISMA-IPD statement?
Section 5: Other considerations
14. Observational studies
Were covariates appropriately conceptualized and categorized as confounders, time-varying confounders, mediators, or other relevant factors?

Abbreviations: IPD, individual-participant data; MA, meta-analysis.

2013, and 2023 to identify observational studies evaluating exposure-outcome relationships in which confounder adjustment would be expected. We excluded articles that were descriptive, predictive, and quasi-experimental (ie, that primarily address confounding through design-based approaches). We randomly selected half of the articles in each journal and publication year for full-text evaluation and identified key study characteristics. We classified the methods reported by each study to select confounders: no confounder adjustment; adjustment but confounders not specified; confounders selected without justification; confounders selected based on an established association with the outcome; confounders selected based on statistical criteria (ie, imbalance between exposure groups, change-in-estimate strategy, or application of a stepwise regression); or confounders selected based on a causal model, either depicted by a DAG or explained in the text. We followed the STROBE reporting guideline for cross-sectional studies.

Results We identified 623 eligible observational studies, including 197 (31.6%) published in medical journals and 426 (68.4%) published in epidemiological journals. Of these, 22 (3.5%) did not report adjusting for confounders, 18 (2.9%) did not specify which confounders were selected, 281 (45.1%) reported selection of confounders without justification, 139 (22.3%) reported selection of confounders based on an established association with the outcome, 121 (19.4%) reported selection of confounders based on statistical criteria, and 42 (6.7%) reported selection of confounders based on a causal model (35 used a DAG and 7 provided an explanation in their text) (**Table 25-1006**). The selection of confounders without justification remained relatively stable between 2003 and 2023 (from 111 of 228 [48.7%] to 68 of 164 [41.5%]), while the use of a causal model to identify confounders increased (from 0 of 228 to 37 of 164 [22.6%]). Differences in the methods used to select confounders were observed across journal type and study design, but not exposure type or funding source.

Conclusions Although our analysis was limited to high-impact medical and epidemiological journals, we found that the reporting of causal models—such as DAGs—has increased over time. However, fewer than one-fourth of studies reported such models in 2023, raising concerns about how confounders are selected and justified in observational research.

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, US, joshua.wallach@emory.edu; [2]Bahiana School of Medicine and Public Health, Salvador, BA, Brazil; [3]Department of Medicine, Yale School of Medicine, New Haven, CT, US.

Conflicts of Interest Joseph S. Ross reported receiving grants from the US Food and Drug Administration; Johnson & Johnson; the Medical Device Innovation Consortium; the Agency for Healthcare Research and Quality; the National Heart, Lung, and Blood Institute; and Arnold Ventures outside the submitted work and is also is an expert witness at the request of relator attorneys, the Greene Law Firm, in a *qui tam* suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc that was settled in September 2022. Dr Ross is a Deputy Editor at *JAMA*. Joshua D. Wallach is supported by Arnold Ventures, Johnson & Johnson through the Yale Open Data Access project, and the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award 1K01AA028258 and previously served as a consultant to Hagens Berman Sobol Shapiro LLP and Dugan Law Firm APLC. Dr Wallach is a Specialty Associated Editor at *JACC*.

Table 25-1006. Methods Used by Observational Studies to Select for Confounders

Year and journal type[a]	Studies, No. (%)					
	No confounder adjustment	Adjustment with confounders not specified	Confounders selected without justification	Confounders selected based on an established association with the outcome	Confounders selected based on statistical criteria[b]	Confounders selected based on causal model
2003						
Medical journal (n = 82)	13 (15.8)	7 (8.5)	37 (45.1)	14 (17.1)	11 (13.4)	0
Epidemiological journal (n = 146)	6 (4.1)	0	74 (50.7)	34 (23.3)	32 (21.9)	0
Total (N = 228)	19 (8.3)	7 (3.1)	111 (48.7)	48 (21.1)	43 (18.9)	0
2013						
Medical journal (n = 70)	0	8 (11.4)	27 (38.6)	14 (20.0)	19 (27.1)	2 (2.8)
Epidemiological journal (n = 161)	3 (1.9)	1 (0.6)	75 (46.6)	47 (29.2)	32 (19.9)	3 (1.9)
Total (N = 231)	3 (1.3)	9 (3.9)	102 (44.2)	61 (26.4)	51 (22.1)	5 (2.2)
2023						
Medical journal (n = 45)	0	2 (4.4)	22 (48.9)	3 (6.7)	9 (20.0)	9 (20.0)
Epidemiological journals (n = 119)	0	0	46 (38.7)	27 (22.7)	18 (15.1)	28 (23.5)
Total (N = 164)	0	2 (1.2)	68 (41.5)	30 (18.2)	27 (16.5)	37 (22.6)
All years and journals (N = 623)	**22 (3.5)**	**18 (2.9)**	**281 (45.1)**	**139 (22.3)**	**121 (19.9)**	**42 (6.7)**

[a]Medical journals: *Annals of Internal Medicine, BMJ, JAMA, Lancet,* and *New England Journal of Medicine*. Epidemiological journals: *American Journal of Epidemiology*, *Annals of Epidemiology*, *Epidemiology*, *European Journal of Epidemiology*, and *International Journal of Epidemiology.*

[b]Imbalance between exposure groups, change-in-estimate strategy, or stepwise regression.

Noninferiority vs Superiority Trials in Cardiovascular Research: Trends, Success in Meeting the Primary Outcome, and Engagement Patterns

Ashkan Hashemi,[1] Isaac Dreyfus,[2] Nicholas Varunok,[3] John Burton,[4] Sina Rashedi,[5] Farbod Zahedi Tajrishi,[6] Behnood Bikdeli[5]

Objective The role of noninferiority trials in cardiovascular research remains unclear. We sought to determine the proportion of noninferiority vs superiority trials, their likelihood of meeting primary outcomes, and their academic impact and online engagement.

Design We reviewed cardiovascular randomized controlled trials (RCTs) published in *The New England Journal of Medicine*, *JAMA*, and *The Lancet* between January 2014 and December 2019. Cardiovascular trials were defined based on review by 2 independent reviewers for topics that included patients with cardiovascular diseases or tested cardiovascular outcomes. Trials were categorized by design (superiority vs noninferiority), and assessed for primary outcome success, academic impact (Google Scholar citations), and online engagement (Altmetric Attention Score). Linear trends were assessed via linear regression. An interaction term was tested between trial design (superiority vs noninferiority) and outcome success (meeting the primary outcome or not) with respect to citation count.

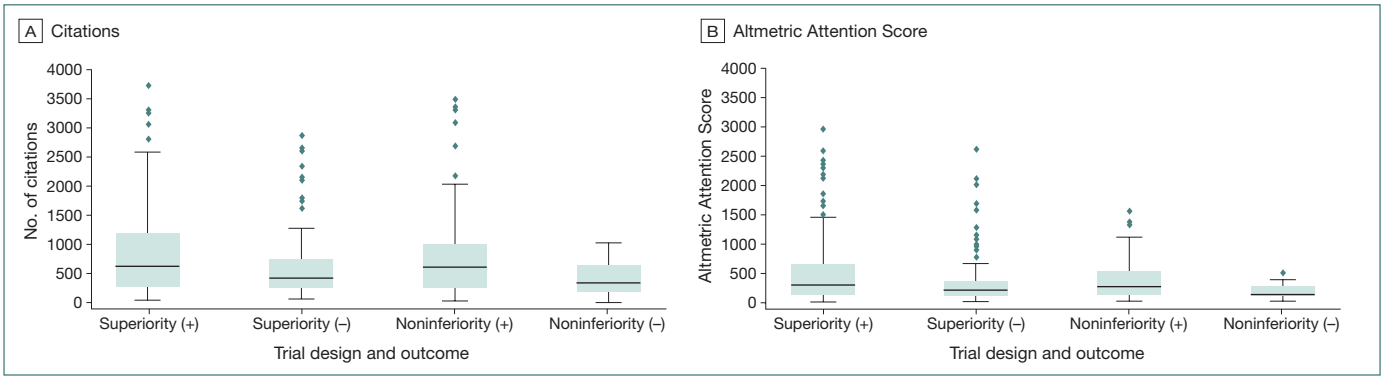
Results Among 429 cardiovascular RCTs, 100 (23.3%) were noninferiority trials, with their proportion increasing over time ($P = .009$). Noninferiority trials were significantly more likely to meet their primary outcome than superiority trials (79% vs 54%; $P < .001$). Median citations were comparable between noninferiority and superiority trials (560 [IQR, 248-851] and 505 [IQR, [274-886], respectively; $P = .83$), as were Altmetric Attention Scores (226 [IQR, 124-481] and 234 [IQR, 120-542], respectively; $P = .70$) (**Figure 25-1150**). Regardless of design, trials that met their primary outcome had higher citation counts than those that did not (superiority trials, 615 [IQR, 280-1200] vs 423 [IQR, 267-734], respectively; $P = .003$; noninferiority trials, 613 [IQR, 261-998] vs 338 [IQR, 201-635], respectively; $P = .049$). There was no significant interaction between trial design and outcome success with respect to citation counts (P for interaction = .87).

Conclusions Noninferiority trials are increasingly shaping cardiovascular research and are more likely to meet their primary outcome, with academic engagement and online attention comparable to superiority trials. The higher success rate observed in noninferiority trials may, in part, reflect the discretionary—and at times overly permissive—approach to noninferiority margin selection, warranting further study. Finally, Altmetric Attention Scores reflect online attention but not clinical impact; future work will assess influence on guidelines and US Food and Drug Administration labeling.

[1]Weill Cornell Medicine, New York, NY, US; [2]David Geffen School of Medicine, University of California, Los Angeles, CA, US; [3]Vanderbilt University Medical Center, Nashville, TN, US; [4]Keck School of Medicine, University of Southern California, Los Angeles, CA, US; [5]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US, bbikdeli@bwh.harvard.edu; [6]Tulane University School of Medicine, New Orleans, LA, US.

Conflict of Interest Disclosures None reported.

Figure 25-1150. Citations and Altmetric Attention Score by Trial Design and Outcome



Citations and Altmetric Attention Score for superiority and noninferiority trials were based on whether they did (+) or did not (–) meet their primary outcomes. Horizontal lines indicate medians, boxes indicate IQRs, and whiskers indicate 95% CIs. Diamonds represent outliers.

Virtual

Estimation of an Upper Limit on the Maximum Effect That Can be Detected in Randomized Trials of Cancer Therapeutics

Benjamin Djulbegovic,[1] Iztok Hozo,[2] Renata Iskander,[5] Austin J. Parish,[3,4] Jonathan Kimmelman,[5] John P. A. Ioannidis[4]

Objective Randomized clinical trials (RCTs) are commonly considered important for detecting small treatment benefits. However, previous studies have shown they are equally likely to detect large ("dramatic") treatment effects.[1-3] Forecasting the likelihood of future large treatment effects is critical for informing policies on resource allocation in conducting human RCTs, including cancer trials.

Design We conducted a systematic review to inform generalized Pareto distribution (GPD) under extreme value theory to predict future maximum treatment effects based on data from the past 65 years. We included consecutive cancer RCTs ("cohorts") identified by funders or trial registries designed to minimize publication bias and analyzed all trials regardless of publication status. Five such cohorts have been described in the literature to date.

Results Between 1955 and 2018, a total of 716 RCTs testing 984 experimental vs standard treatments in 349,947 patients were conducted and published by 2022, averaging approximately 20 RCTs per year. The shape parameter of the GPD had positive values, indicating no upper limit on maximum treatment effects. We found that the treatment with the largest effect in the past had an odds ratio (OR) of 45 (95% CI, 2-1008). If effect patterns remain the same and a similar pace of performing RCTs continues, the largest predicted future effect over the next 50 years would be an OR of 23 (95% CI, 4-43) (**Figure 25-0942**). We estimated a 20% probability of detecting new treatment effects with an OR greater than 50 in the next 50 years. We also found that conducting more RCTs (40 or 60 per year) would double or triple the probability of detecting breakthrough treatments with dramatic effects.

Conclusions Our analysis indicates there may be no upper bound on the maximum discoverable treatment effects in cancer RCTs, but effect estimates will likely remain within the range of those observed between 1955 and 2022. Conducting more RCTs would accelerate the detection of treatments with large effects.

References

1. Djulbegovic B, Kumar A, Glasziou P, Miladinovic B, Chalmers I. Medical research: Trial unpredictability yields predictable therapy gains. *Nature*. 2013;500(7463):395-396. doi:10.1038/500395a

2. Hozo I, Djulbegovic B, Parish AJ, Ioannidis JPA. Identification of threshold for large (dramatic) effects that would obviate randomized trials is not possible. *J Clin Epidemiol*. 2022;145:101-111. doi:10.1016/j.jclinepi.2022.01.016

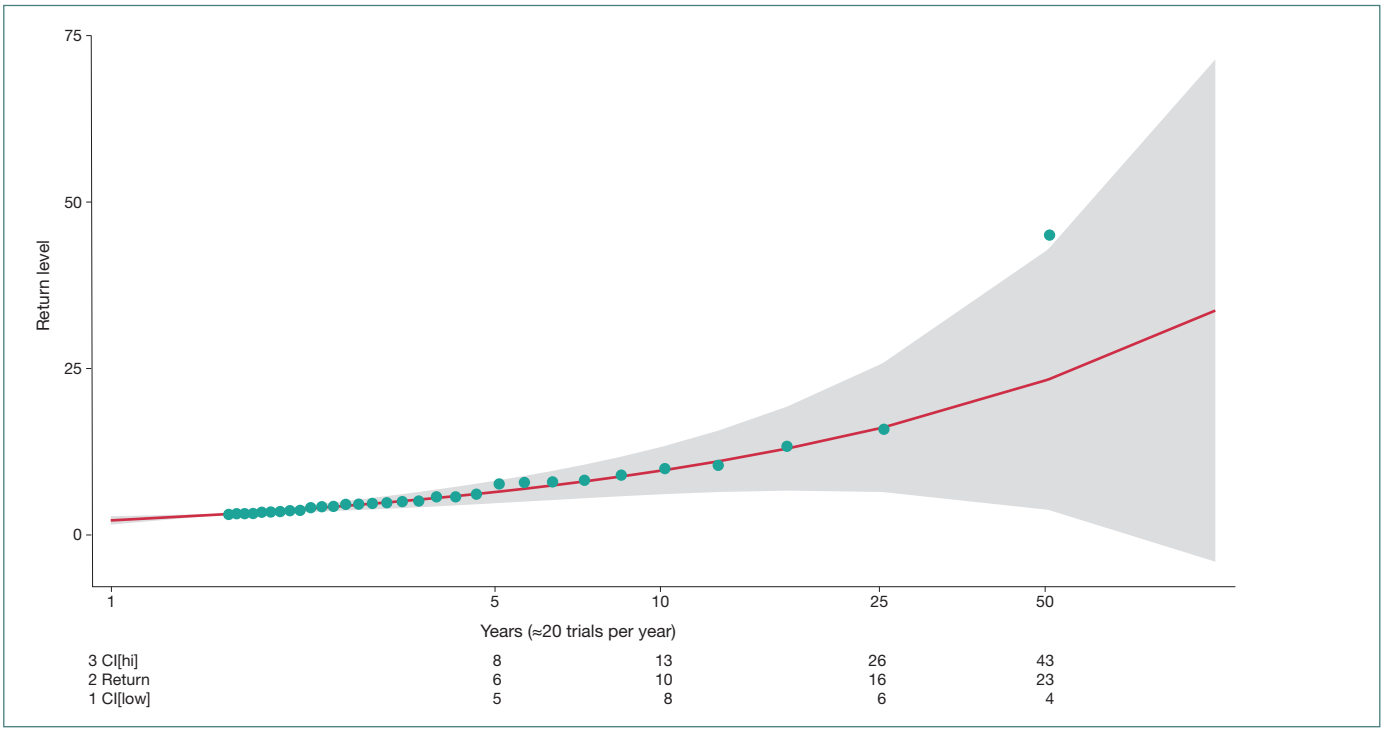
3. Djulbegovic B, Kumar A, Soares HP, et al. Treatment success in cancer: new cancer treatment successes identified in phase 3 randomized controlled trials conducted by the National Cancer Institute-sponsored cooperative oncology groups, 1955 to 2006. *Arch Intern Med*. 2008;168(6):632-642. doi:10.1001/archinte.168.6.632

[1]Medical University of South Carolina, Division of Medical Hematology and Oncology, Department of Medicine, Charleston, SC, US, djulbegov@musc.edu; [2]Department of Mathematics, Indiana University Northwest, Gary, IN, US; [3]Department of Emergency Medicine, Lincoln Medical Center, Bronx, NY; [4]Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine; Department of Epidemiology and Population Health, Stanford University School of Medicine; Department of Biomedical Data Science, Stanford University School of Medicine; Department of Statistics, Stanford University School of Humanities and Sciences, Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, US; [5] Department of Equity, Ethics and Policy School of Population and Global Health, McGill University, Montreal, QC, Canada.

Conflict of Interest Disclosures John P. A. Ioannidis 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.

Additional Information Datasets for this work were obtained with support of grants from the US National Institute of Health: R01CA140408, R01NS044417, R01NS052956, and R01CA133594 (Benjamin Djulbegovic).

Figure 25-0942. Predicted Future Effect



Return level represents the magnitude of the treatment effect that is expected to be exceeded, on average, once every N years (where N is the return period). The maximum observable treatment effect over the next 50 years is equal to odds ratio of 23 (95% CI, 4-43).

Performance and Practicality of a Randomized Clinical Trial Classifier in Systematic Literature Reviews vs a Traditional Approach

Ambar Khan,[1] Ania Bobrowska,[2] Chloe Coelho,[2] Hannah Frost,[2] Swati Kumar,[2] Hannah Russell,[3] Anna Noel-Storr,[4] Molly Murton[2]

Objective This study aimed to evaluate a traditional (validated search filter) vs machine learning (classifier) approach for identifying randomized clinical trials (RCTs), in the context of a real-life systematic literature review (SLR) project.

Design A real-life SLR was conducted following standard practices as recommended by the Cochrane Handbook,[1] covering all identification stages from electronic database searches to full-text review, to identify RCTs on fibrodysplasia ossificans progressiva (FOP; a rare disease). Two approaches were compared for the searches conducted in January 2025: (1) combining population search terms with the Scottish Intercollegiate Guidelines Network (SIGN) RCT search filter ("filter" approach) and (2) using population search terms only and running the results through the Cochrane RCT classifier at a 99% recall threshold ("classifier" approach). All subsequent stages were conducted in an identical manner by the same review team, with both abstracts and full texts dual-screened by independent reviewers against prespecified eligibility criteria. The time taken for each approach was measured and reviewer feedback was gathered via a modified NASA Task Load Index. Performance metrics for each approach were calculated, including the accuracy, sensitivity, specificity, and precision. McNemar test was used to assess statistical significance of the differences in the performance metrices, and a *P* value <.05 was considered statistically significant.

Results Prior to application of the classifier or filter, 9588 records were found for the FOP population. After application of the classifier or filter, 2327 and 1582 abstracts were identified as RCTs, respectively. At the end of the full-text review stage for each approach, the same 10 papers reporting on FOP RCTs were identified, demonstrating equal efficacy in terms of finding relevant papers in the context of a real-life SLR. However, when looking more generally at the classification of RCT vs non-RCT publications within the 9588 records (irrespective of ultimate inclusion in the SLR), sensitivity was significantly higher for the classifier vs the filter (McNemar test $P < .001$; **Table 25-0997**). The trade-off for higher sensitivity of the classifier was that it had a significantly lower specificity and therefore required 31% more time than the filter approach (25.79 vs 17.86 hours). This also resulted in a less positive user experience in effort and frustration domains.

Conclusions For SLRs, where the identification of all relevant evidence is crucial, the performance of the classifier was superior to a traditional search filter. However, this was at the cost of increased time and reduced specificity. Further testing of the application of machine learning classifiers to augment literature review processes in real-life projects (including other, larger disease areas) is required. The risk of bias in this research, stemming from the researchers being unblinded to the approach, should also be noted.

Table 25-0997. Accuracy, Sensitivity, Specificity, and Precision of the Classifier and Filter Approaches for Identifying Any Randomized Clinical Trials (N = 9588)

Performance metric	Calculation[a]	Classifier approach, %	Filter approach, %	McNemar test *P* value
Accuracy	(TP+TN)/(TP+TN+FP+FN)	80.8	86.7	<.001
Sensitivity	TP/(TP+FN)	98.6	82.1	<.001[b]
Specificity	TN/(TN+FP)	79.8	86.9	<.001
Precision	TP/(TP+FP)	21.1	25.6	Not calculable[c]

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

[a]The raw values used in the calculations were as follows: classifier: TP = 491, TN = 7254, FP = 1836, and FN = 7; filter: TP = 409, TN = 7900, FP = 1190, and FN = 89.

[b]The classifier approach performs significantly better than the filter approach.

[c]Because there were no discordant pairs, the McNemar test could not be performed.

Reference

1. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3*. Cochrane; 2022.

[1]Costello Medical, London, UK; [2]Costello Medical, Cambridge, UK, molly.murton@costellomedical.com; [3]Costello Medical, Manchester, UK; [4]Cochrane, London, UK.

Conflict of Interest Disclosures Ambar Khan, Ania Bobrowska, Chloe Coelho, Hannah Frost, Swati Kumar, Hannah Russell, and Molly Murton reported being employees of Costello Medical. No other disclosures were reported.

Funding/Support This study was funded by Costello Medical.

Role of the Funder/Sponsor Costello Medical supported all aspects of the study, including 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.

Retractions

In-person

An Audit and Feedback Intervention to Reduce Inappropriate Citation of Retracted Literature in the Pain and Anesthesiology Fields

Michael C. Ferraro,[1,2] Aidan G. Cashin,[1,2] Amanda C. de C. Williams,[3] Emma Fisher,[4] Gavin Stewart,[5] Christopher Eccleston,[4,6,7] Neil E. O'Connell[8]

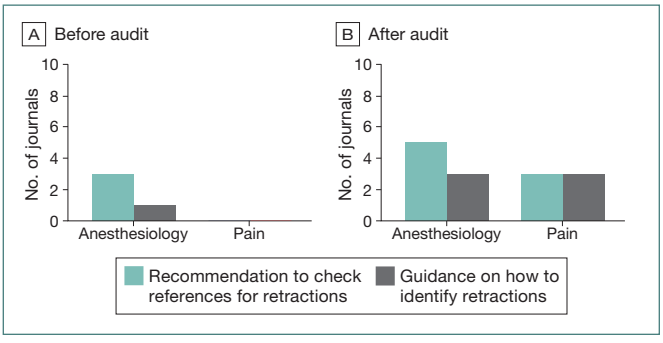
Objective Continued citation of retracted literature is a threat to scientific integrity. Most authors who cite retractions are unaware the article has been withdrawn.[1] While the International Committee of Medical Journal Editors (ICMJE) requires authors to check that no references in a manuscript cite retracted articles,[2] few journals include this requirement in author submission guidelines.[3] This study evaluated the effect of an audit and feedback intervention on retraction checking instructions in the top 20 pain and top 20 anesthesiology journal author submission guidelines.

Design This pre-post audit and feedback intervention evaluated author submission guidelines of the top 20 pain and top 20 anesthesiology journals (according to Journal Citation Reports 2022 Impact Factor) for (1) recommendations to check manuscript references for retractions and (2) guidance on how to identify retractions. Each item was judged as yes or no (1 or 0, respectively) and given a total score out of 2. The feedback intervention was targeted at the journal editors in chief via a single email contact. Based on the journal's score, a tailored letter was constructed (for pain and anesthesiology journals separately) including the journal score, comparison with other journals evaluated, and guidance (including a template) to implement ICMJE recommendations within author guidelines. A follow-up audit was performed at 6 months. This study was approved by University of New South Wales ethics and registered on Open Science Framework on April 2, 2024.

Results The baseline audit was performed in April 2024. The mean (SD) impact factor was 3.6 (1.5) for pain journals and 4.6 (2.6) for anesthesiology journals. Twelve of 20 pain journals and 9 of 20 anesthesiology journals were ICMJE signatories. At baseline, 0 of 20 pain journals and 3 of 20 anesthesiology journals included recommendations to check references for retractions and 0 of 20 pain journals and 1 of 20 anesthesiology journals provided guidance on how to identify retractions. At 6-month follow-up in October 2024, 3 of 20 pain journals and 5 of 20 anesthesiology journals included recommendations to check references for retractions and 3 of 20 pain journals and 3 of 20 anesthesiology journals provided guidance on how to identify retractions (**Figure 25-0881**). Overall, this represented an absolute score change of 25%. Correspondence with editors of 2 pain journals and incorporation of the guidance template in 1 pain journal confirmed that these changes resulted from the feedback intervention.

Conclusions The number of pain and anesthesiology journals with author instructions to check references for retractions increased following a tailored audit and feedback intervention. However, overall uptake of the feedback was low, suggesting that more sustained editorial engagement strategies may be necessary to facilitate implementation of recommendations. Future studies should evaluate whether implementation of retraction screening recommendations into author submission guidelines reduces citation of retracted literature.

Figure 25-0881



References

1. De Cassai A, Geraldini F, De Pinto S, et al. Inappropriate citation of retracted articles in anesthesiology and intensive care medicine publications. *Anesthesiology*. 2022;137(3):341-350. doi:10.1097/ALN.0000000000004302

2. International Committee of Medical Journal Editors. Preparing a manuscript for submission to a medical journal. Accessed January 15, 2025. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html

3. Boudry C, Howard K, Mouriaux F. Poor visibility of retracted articles: a problem that should no longer be ignored. *BMJ*. 2023;381:e072929. doi:10.1136/bmj-2022-072929

[1]Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, New South Wales, Australia, m.ferraro@neura.edu.au; [2]School of Health Sciences, Faculty of Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia; [3]Research Department of Clinical, Educational and Health Psychology, University College London, London, UK; [4]Centre for Pain Research, University of Bath, Bath, UK; [5]School of Natural and Environmental Sciences, University of Newcastle upon Tyne, Newcastle, UK; [6]Department of Health and Clinical Psychology, the University of Ghent, Ghent, Belgium; [7]Department of Psychology, The University of Helsinki, Helsinki, Finland; [8]Department of Health Sciences, Centre for Wellbeing Across the Lifecourse, Brunel University of London, Uxbridge, UK.

Conflict of Interest Disclosures Aidan G. Cashin, Amanda C. de C. Williams, Emma Fisher, and Christopher Eccleston are editorial board members for 2 of the pain journals (*PAIN*, *Journal of Pain*) included in this audit. No other disclosures were reported.

Funding/Support No specific funding was received for this project. Michael C. Ferraro was supported by an Australian Government Research Training Program PhD scholarship and a Neuroscience Research Australia top-up scholarship.

Role of the Funder/Sponsor The sponsor, Neuroscience Research Australia, had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and decision to submit the manuscript.

Additional Information This study was registered on Open Science Framework at https://osf.io/gbkjy/.

Prevalence of and Reasons for Retractions of Traditional Chinese Medicine Research Publications by Authors From Mainland China in International Peer-Reviewed Journals

Jing Cui,[1] Nan Yang,[1] Kexin Ji,[1] Dingran Yin,[1] Chen Shen,[2] Zhaolan Liu,[2] Han Tan,[3] Yaxin Sun,[4] Zhaoqi Huo,[5] Shuo Liu,[5] Huiyu Wang,[5] Xintong Zhang,[6] Jing Guo,[1] Yufei Wang,[7] Xiaoqi Ren,[5] Vincent C. H. Chung,[8] Jianping Liu[2]

Objective Publication retractions, especially in the field of medicine, have long been a matter of significant concern.[1] The trends and characteristics of retractions of publications on Traditional Chinese medicine (TCM) by authors from mainland China remain unclear. This study is dedicated to exploring the prevalence and underlying reasons of retractions in TCM publication.

Design This cross-sectional study identified and analyzed publications that were marked as retractions or withdrawal. It encompassed retractions of TCM literature published by mainland Chinese authors in international peer-reviewed journals from January 2000 to November 2024. Seven medical databases were systematically searched, including 3 English-language databases (PubMed, Embase, and Cochrane Library) and 4 Chinese-language databases (China National Knowledge Infrastructure, Wanfang Data, VIP Database, SinoMed), as well as the Retraction Watch Database. Duplicate data were excluded. Article screening and data extraction were independently conducted by 2 authors. Discrepancies were resolved through discussion. The retractions of TCM publications by mainland Chinese authors were thoroughly analyzed. The numbers and characters of retractions across different study types were compared, and retraction reasons were comprehensively summarized according to Committee on Publication Ethics (COPE) guidelines. Main outcomes included the prevalence, reasons, and relevant factors of retraction (year, authors' region and affiliation, study types, study subjects, and reasons).

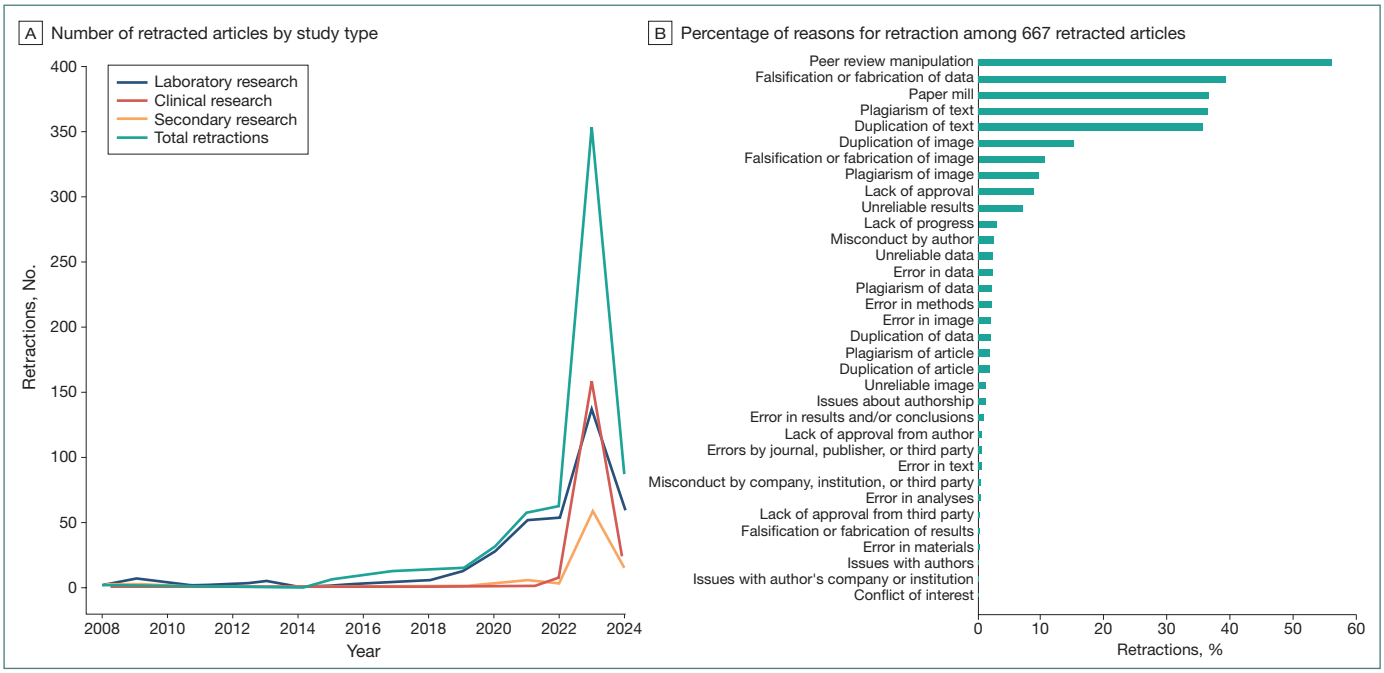
Results A total of 24,479 retracted articles were initially identified, and 679 retracted publications on TCM met inclusion criteria. Retractions encompassed laboratory studies (381 [56.1%]), clinical studies (189 [27.8%]), and secondary studies (109 [16.1%]). Notably, the number of retractions surged to 354 articles (52.1%) in 2023 (**Figure 25-0930**). Most first authors were affiliated with Grade A tertiary hospitals (the highest accreditation level in China's health care system; 415 [61.1%]). Active-ingredient extracts of TCM (289 [42.6%]) and neoplasms (145 [22.6%]) emerged as the most prevalent study subjects. A total of 77.6% of retractions were attributed to multiple reasons, so that the sum of the percentage of reasons may exceed 100%. A total of 12 retracted articles had no specific reasons found. Author misconduct drove 576 retractions (86.4%), predominantly involving peer review manipulation (374 [56.1%]), falsification or fabrication of data (262 [39.3%]), and paper mills (244 [36.6%]).

Conclusions Currently, retractions of TCM publications have become increasingly conspicuous, particularly those from higher-level academic institutions and hospitals. Systemic author misconduct, notably compromised peer review processes and industrialized fraud, may constitute the primary driver. However, some publishers regularly remove retracted literature from the platform, which results in omissions in the data. To promote scientific credibility in TCM publications, we recommend (1) enhanced institutional accountability mechanisms, (2) standardized retraction policies aligned with COPE guidelines, (3) mandatory research integrity training, and (4) transparent postretraction investigations.[2]

References

1. Else H. Biomedical paper retractions have quadrupled in 20 years—why? *Nature*. 2024;630(8016):280-281. doi:10.1038/d41586-024-01609-0

Figure 25-0930. Retracted Traditional Chinese Medicine Articles



A, We analyzed 665 retractions by year of retraction. A total of 14 eligible retractions could not be analyzed for retraction time because they had already been removed from the platform during our data extraction. B, A total of 5 laboratory studies, 1 clinical study, and 6 secondary studies did not report specific reasons for retraction. A total of 77.6% of retractions were attributed to multiple reasons; the sum of percentages of reasons for retraction may exceed 100%.

2. Bakker CJ, Reardon EE, Brown SJ, et al. Identification of retracted publications and completeness of retraction notices in public health. *J Clin Epidemiol*. 2024;173:111427. doi:10.1016/j.jclinepi.2024.111427

[1]Qi-Huang Chinese Medicine School, Beijing University of Chinese Medicine, Beijing, China; [2]Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China, liujp@bucm.edu.cn; [3]School of Acupuncture-Moxibustion and Tuina, Beijing University of Chinese Medicine, Beijing, China; [4]School of Management, Beijing University of Chinese Medicine, Beijing, China; [5]School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China; [6]School of Humanities, Beijing University of Chinese Medicine, Beijing, China; [7]College of Continuing Education, Hebei University of Chinese Medicine, Shijiazhuang, China; [8]Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China.

Conflict of Interest Disclosures None reported.

Funding/Support High-level Traditional Chinese medicine key subjects construction project of National Administration of Traditional Chinese Medicine—Evidence-Based Traditional Chinese Medicine (award zyyzdxk-2023249).

Role of the Funder/Sponsor The sponsor contributed to the decision to submit the abstract for presentation.

Additional Information Jing Cui and Nan Yang contributed equally to this work, and Kexin Ji and Dingran Yin contributed equally to this work. Vincent C. H. Chung is a co–corresponding author (vchung@cuhk.edu.hk).

Citation Context Analysis of Retracted Articles: Leveraging Retraction Reasons to Track Unreliability in Citing Literature

Yagmur Ozturk,[1] Frédérique Bordignon,[2,3] Cyril Labbé,[1] François Portet[1]

Objective To address the challenge of tracking the use of unreliable sources in scientific literature, we analyzed citations to retracted articles (CRAs) using retraction reasons (RRs) and citation context (CC) analysis. Our aim was to inform the development of automated tools for postpublication peer review (PPPR) by identifying instances in which a CRA is not only an unreliable source, but may also introduce unreliability into the citing work itself, depending on the way it is cited. We examined the degree to which the content of the citation context aligns with the RR, indicating potential risk for the integrity of the citing article.

Design RRs were obtained from the Retraction Watch database[1] for a set of 22,558 publications that cite retracted sources accessed from the Problematic Paper Screener.[2] We selected 29 of 108 retraction reasons that directly relate to scientific content of the article (eg, unreliable or fabricated data or results). To ensure access to CCs, we limited our dataset to open access citing articles and excluded those that had themselves been retracted. This yielded 4979 citing articles and 7424 unique CRA instances; 88% of these were preretraction citations. We then limited the rest of the analysis to a subset of 3049 CCs from the PubMed Central Open Access corpus.[3] We used regular expressions to search for terms referencing concepts related to the retraction reasons (eg, “data,” “results,” “findings”) and manually validated their use in the CCs.

Results Of 3049 CCs, we identified 57 that explicitly mentioned the exact issue mentioned in the RR. Among these, 12 emphasized similarity or consistency with the retracted findings, which are instances we interpret as potential indicators of unreliability in the citing paper, meriting PPPR. These 57 CCs span 18 research fields and involve 60 distinct RRs (some overlapping with the selected 29). The cases are distributed across a range of publishers and fields, offering a basis to be used in the development of automatic systems.

Conclusions While retracted publications are unreliable sources, we argue that how the retracted work is used and why it was retracted matter more than the fact that a retracted reference is found in a publication. If the retracted source is used to support the citing work, this calls for a secondary assessment when it is detected. By combining RRs with CC analysis, our approach informs a practical method to flag such cases for PPPR and supports the development of tools for large-scale CRA assessment.

References

1. The Center for Scientific Integrity. Retraction Watch Database. https://retractiondatabase.org/

2. Cabanac G, Labbé C, Magazinov A. The 'Problematic Paper Screener' automatically selects suspect publications for post-publication (re)assessment. *arXiv*. Preprint posted online October 7, 2022. doi:10.48550/arXiv.2210.04895

3. National Library of Medicine. PMC Open Access Subset. Accessed January 30, 2025. https://pmc.ncbi.nlm.nih.gov/tools/openftlist/

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France, yagmur.ozturk@univ-grenoble-alpes.fr; [2]Ecole Nationale des Ponts et Chaussées, Institut Polytechnique de Paris, Marne-la-Vallée, France; [3]LISIS, INRAE, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France.

Conflict of Interest Disclosures None reported.

Funding/Support We acknowledge the NanoBubbles project that has received Synergy grant funding from the European Research Council within the European Union's Horizon 2020 program (grant agreement number 951393) (https://cordis.europa.eu/project/id/951393).

Evaluating Approaches for Identifying Retracted Articles and Retraction Notices in Systematic Review Searching

Caitlin J. Bakker,[1,2] Erin E. Reardon,[3] Nicole Theis-Mahon,[4] Sara Schroter,[5,6] Lex Bouter,[7,8] Maurice P. Zeegers[2]

Objective Systematic reviews gather, appraise, and synthesize studies to inform research, practice, and policy. However, the inclusion of retracted articles, which often contain flaws and falsified or fabricated data, undermines the credibility of systematic reviews. Identifying retracted articles is challenging, as they are inconsistently flagged.[1,2] Our study validated and compared approaches to identify retracted articles and their retraction notices.

Design Our study, guided by an advisory panel of information specialists and researchers, evaluated approaches for identifying retracted publications from 8 health sciences databases (Cochrane Library, Embase.com, Ovid Embase, Ovid Medline, Ovid PsycINFO, PubMed, Scopus, and Web of Science). Using a reference set of 43,544 retracted publications and 27,755 associated retraction notices from Retraction Watch, we identified items found in each database. From August 10 to 14, 2024, we determined how many of the available items could be retrieved using each approach per database. Two search strategies, database indexing, and 2 citation managers were evaluated. The complete methodology, including search strategies, is available in our protocol.[3] Recall (sensitivity) was calculated to evaluate identification effectiveness.

Results Recall of retracted publications and notices varied across databases and retrieval approaches. Across databases, search strategy 2 consistently achieved the highest recall, with values ranging from 74.6% to 96.9%. Search strategy 1 also performed strongly, particularly in PubMed and Web of Science (both >93%). In contrast, indexing-based retrieval showed variable performance, with high recall in PubMed (94.6%) and Ovid Medline (94.5%) but much lower in Embase.com (40.6%) and PsycINFO (34%). Citation manager tools (EndNote and Zotero) yielded lower recall, with values rarely exceeding 64%. Recall was lowest for Ovid Embase and PsycINFO regardless of method, while PubMed and Web of Science showed the highest recall.

Conclusions There was substantial variability in the ability of databases and retrieval approaches to identify retracted publications and notices. No single approach captured all items, underscoring the need for multiple approaches in an iterative identification process.

References

1. Bakker CJ, Reardon EE, Brown SJ, et al. Identification of retracted publications and completeness of retraction notices in public health. *J Clin Epidemiol*. 2024;173:111427. doi:10.1016/j.jclinepi.2024.111427

2. Boudry C, Howard K, Mouriaux F. Poor visibility of retracted articles: a problem that should no longer be ignored. *BMJ*. 2023;381:e072929. doi:10.1136/bmj-2022-072929

3. Bakker C, Reardon EE, Theis-Mahon NR, Schroter S, Bouter L, Zeegers M. Validation and comparison of methods to identify retracted publications during information retrieval. Open Science Framework. Cited June 7, 2025. https://osf.io/rwzym/

[1]University of Regina, Regina, SK, Canada, caitlin.bakker@uregina.ca; [2]Maastricht University, Maastricht, the Netherlands; [3]Emory University, Atlanta, GA, US; [4]University of Minnesota, Minneapolis, US; [5]BMJ, London, UK; [6]London School of Hygiene and Tropical Medicine, London, UK; [7]Amsterdam University Medical Center, Amsterdam, the Netherlands; [8]Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

Conflict of Interest Disclosures Caitlin J. Bakker is cochair of the National Information Standards Organization Communication

of Retractions, Removals and Expressions of Concern Standing Committee. No other disclosures were reported.

Funding/Support This research is part of an ongoing PhD collaboration between *BMJ* and the team Meta-Research at Maastricht University (UM) on the responsible conduct of publishing scientific research. *BMJ* is published by BMJ Group, a wholly owned subsidiary of the British Medical Association. UM is a public legal entity in the Netherlands. This study is part of Caitlin J. Bakker's self-funded BMJ/UM PhD. No exchange of funds has taken place for this research project.

Role of the Funder/Sponsor The authors are wholly responsible for the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the abstract; and decision to submit the abstract for presentation.

Additional Information All authors express their own opinions and not necessarily that of their employers.

Postretraction References in Biomedical and Clinical Sciences

Guillaume Cabanac,[1,2] Kathryn Weber-Boer[3]

Objective We analyzed postretraction citations of retracted publications in biomedical and clinical sciences (BCS). Previous studies have investigated the citation of retracted clinical trials[1] and the citation of retracted publications in policy documents,[2] but the share of references cited after their retraction in BCS literature is unknown. We identified the articles citing postretracted references and present the results aggregated at the journal and publisher levels.

Design We used Dimensions data accessed in BigQuery to delineate the corpus, including publications of the type "article" classified as "review article" or "research article," with the field of research category BCS (field 32 in the ANZSRC 2020 classification), published between 2005 and 2024. These 12,229,418 articles were published in 43,767 journals from 11,613 publishers and contained 28,037,343 references. We extracted the year of publication, journal title, publisher name, and list of references cited. We retained those references that were removed, retracted, or withdrawn (henceforth, retracted), based on data derived from Crossref, PubMed, Retraction Watch, and retraction notices issued by publishers. We excluded publications that cited retractions with a retraction year earlier than or the same as the year of publication to consider only postretraction references.

Results We observed a growing number of publications in the BSC literature that cited retracted publications. In 2005, there were 4000 publications referencing 1074 retracted publications 4755 times. This grew to 16,599 publications referencing 7631 retracted publications 18,614 times in 2024, a 415% increase in publications, 710% increase in retracted publications cited, and 391% increase in retracted citations. This compared with an increase over the same period of 267% for publications in BSC (from 368,338 to 983,837) and 390% in citations (8,495,152 to 33,143,231). This is an increase in the percentage of publications with retracted references of 1.08% to 1.69% of all BSC publications. The number of retracted publications referenced doubled since 2021. The connection between journal and publisher reflected the ownership relationship as of approximately January 2025. The journals in which postretraction referencing of retracted literature occurred are many (7225 distinct journals [17%]), but half the publications (35,394 of 69,719) are currently owned by 4 major publishers: Elsevier (14,589 [21%]), Springer Nature (11,201 [16%]), Wiley (5987 [9%]), and Frontiers (3617 [5%]) (**Table 25-1151**).

Conclusions The publishers of articles responsible for the most citations of postretracted references are the largest publishers of scientific literature by any measure. We could have ranked or displayed the results by publisher or journal as a percentage of publications overall; however, in this case, quantity matters more than share. This practice is preventable, with system-wide corrective efforts by the publishing industry, such as prepublication checks for retracted references, as a complement to the postpublication checks of retracted references that are also needed.[3]

References

1. Kataoka Y, Banno M, Tsujimoto Y, et al. Retracted randomized controlled trials were cited and not corrected in systematic reviews and clinical practice guidelines. *J Clin Epidemiol*. 2022;150:90-97. doi:10.1016/j.jclinepi.2022.06.015

2. Malkov D, Yaqub O, Siepel J. The spread of retracted research into policy literature. *Quant Sci Stud*. 2023;4(1):68-90. doi:10.1162/qss_a_00243

3. Cabanac G. Chain retraction: how to stop bad science propagating through the literature. *Nature*. 2024;632(8027):977-979. doi:10.1038/d41586-024-02747-1

[1]Université de Toulouse, IRIT (UMR 5505 CNRS), Toulouse, France; [2]Institut Universitaire de France, Paris, France; [3]Digital Science, London, UK, k.weberboer@digital-science.com.

Conflict of Interest Disclosures Kathryn Weber-Boer is employed by Digital Science, which provided access to the Dimensions datasets that this work analyzed. Guillaume Cabanac received funding from the Institut Universitaire de France.

Additional Information Guillaume Cabanac and Kathryn Weber-Boer are co–corresponding authors.

Retracted Publications Referenced in Clinical Guidelines

Lonni Besançon,[1] Guillaume Cabanac,[2,3] Kathryn Weber-Boer[4]

Objective We identified retracted references in clinical guidelines published between 1991 and 2025 that were cited before and after retraction to study the scale of this contamination. Clinical guidelines have a concrete impact on the practice of medicine and, while active, are meant to reflect the state of the art. Previous work[1] has explored the citation of retracted trials in systematic reviews and clinical guidelines, showing both pre- and postretraction citations. We explored a population of retracted publications cited in clinical guidelines.

Table 25-1151. Journals and Publishers With the Most Frequent Publications With Retracted References

Publisher	Journal	2005-2014, No.		2015-2024, No.		Total No.	
		Publications	Retracted references	Publications	Retracted references	Publications	Retracted references
Other publishers	Other journals	6355	3679	28,010	17,459	34,365	*21,138*
	PLOS One (Public Library of Science)	631	529	964	876	1595	*1405*
	Oncology Letters	29	30	490	502	519	*532*
	Oncotarget	33	31	622	485	655	516
Elsevier	Other journals	5378	3363	17,054	11,937	22,432	*15,300*
	Biomedicine & Pharmacotherapy	8	8	865	962	873	*970*
	Heliyon			408	447	408	*447*
	Medicine	8	7	6	4	14	11
Springer Nature	Other journals	3208	2204	13,294	9840	16,502	*12,044*
	Scientific Reports	18	17	1208	1118	1226	1135
Wiley	Other journals	2211	1746	7360	6222	9571	7968
Frontiers	Other journals	118	111	3534	3061	3652	*3172*
	Frontiers in Immunology	23	20	810	774	833	*794*
	Frontiers in Oncology	20	20	897	900	917	*920*
	Frontiers in Pharmacology	7	7	943	992	950	999
MDPI	Other journals	71	66	4011	3387	4082	*3453*
	Cancers	19	19	840	833	859	*852*
	Nutrients	17	19	653	473	670	492
Wolters Kluwer	Other journals	1296	1084	3705	3056	5001	*4140*
	Medicine	2	3	548	547	550	550
Sage Publications	Other journals	483	453	1941	1736	2424	2189
Oxford University Press	Other journals	830	719	1576	1429	2406	*2148*

Design We drew on 2 new datasets: clinical guidelines from Altmetric (n = 20,553) and retractions identified by Dimensions (n = 63,047). The clinical guidelines dataset[2] (articles indexed in Dimensions classified as clinical guidelines and documents published online by learned societies and associations) should consist of the latest version of each guideline, although as guidelines are updated, the guidelines that we studied may become obsolete. This dataset included links to guidelines and a list of publications referenced in each guideline. The retractions dataset was derived from CrossRef, PubMed, and Retraction Watch, and by matching publisher notices to publications; it included an identifier for the retracted publication and relevant linked notices with their type and date. We identified publications found both in the list of guideline references and in the list of retracted publications and selected those cited after their retraction date (398 unique guideline-retracted publication pairs). We manually examined a randomly selected 10% of the accessible English-language publications to assess the accuracy of the dataset (42 guideline-reference pairs). Our evaluation of this sample of our dataset assessed both whether the retraction identification was accurate and whether the citation of these publications in the guidelines was critical (or mentioned the retraction).

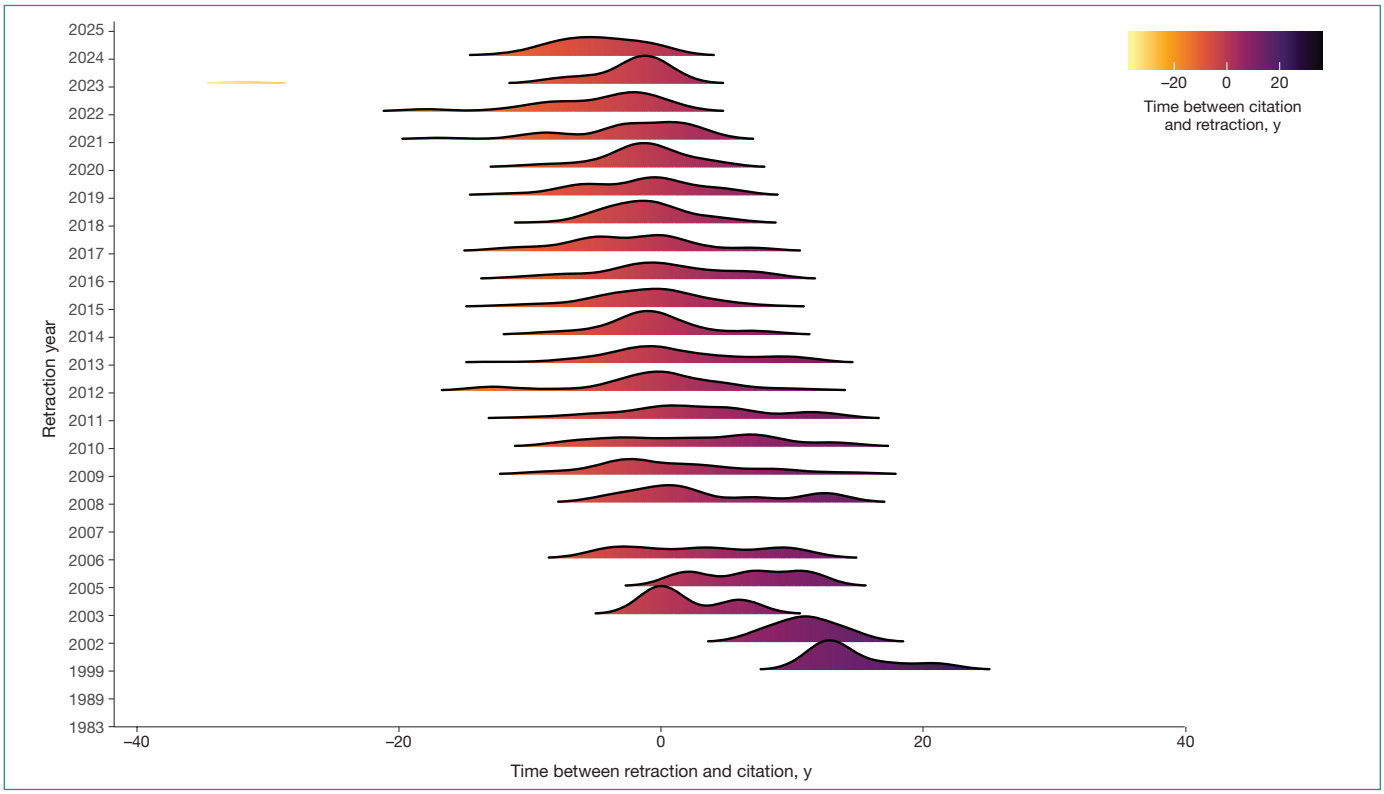
Results We considered publications cited in the year of retraction to be preretraction citations. We found 487 guidelines containing 596 references to 319 publications that had been retracted after citation. Of these, 179 guidelines (37%) referenced 134 retracted publications 204 times within 5 years after retraction (**Figure 25-1004**). There were 107 guidelines (22%) containing 126 references to 73 publications that had been retracted more than 5 years before citation. All retraction datasets, including ours, contained inaccuracies. We could not find retraction notices for 4 references marked as retracted. Only 2 retracted references were cited critically (presented with some doubt as to the strength of their methods or conclusions), and none acknowledged the retracted status of the reference. Most references were used either as background literature or to support claims and recommendations made in the guidelines.

Conclusions Postretraction citations are preventable. The longer after the retraction appears, the more concerning the citation. In guidelines in particular, the persistence of retracted references is concerning. Our results highlight the need to systematically screen both new and existing clinical guidelines for contamination.

References

1. Kataoka Y, Banno M, Tsujimoto Y, et al. Retracted randomized controlled trials were cited and not corrected in systematic reviews and clinical practice guidelines. *J Clin Epidemiol*. 2022;150:90-97. doi:10.1016/j.jclinepi.2022.06.015

Figure 25-1004. Time Between Retractions and Citations in Clinical Guidelines



2. Altmetric. Clinical guidelines. Accessed June 19, 2025. https://www.altmetric.com/about-us/our-data/clinical-guidelines/

[1]Linköping University, Linköping, Sweden; [2]Université de Toulouse, Institut de Recherche en Informatique de Toulouse (UMR 5505), Centre National de la Recherche Scientifique, Toulouse, France; [3]Institut Universitaire de France, Paris, France; [4]Digital Science, London, UK, k.weberboer@digital-science.com.

Conflict of Interest Disclosures Kathryn Weber-Boer is employed by Digital Science, which provided access to the Altmetric and Dimensions datasets that this work analyzed. Lonni Besançon has received a catalyst grant from Digital Science. No other disclosures were reported.

Retraction of Systematic Reviews and Clinical Practice Guidelines

Ivan D. Florez,[1,2] Alberto Henriquez,[3,4] Andrés F. Estupinan-Bohorquez[3,5]

Objective We described the influence of retracted systematic reviews and meta-analyses (SRMAs) on clinical practice guidelines (CPGs) and the characteristics of retractions in CPGs.

Design This cross-sectional study was conducted in 2 stages based on searches focused on the Retraction Watch (RW) database and MEDLINE from inception to November 30, 2024. In the first stage, we included SRMAs. We described the reasons for retractions, and we recategorized them based on our assessment. We identified the CPGs that cited the SRMA in the Google Scholar database. In the second stage, we included the retracted CPGs, described the reasons for retractions, and recategorized them into ethical or nonethical reasons based on our assessment. Nonethical reasons were categorized as editorial or administrative or outdated guidelines, while ethical reasons were reported according to RW categories. We used descriptive statistics to summarize the findings.

Results In the first stage, we included 377 SRMAs, of which 211 (56.0%) were retracted due to peer review or publication manipulation (eg, detected "fake" reviewers); 30 (8.0%), due to duplicate or redundant publication; and 136 (36.1%), due to intellectual or authorship disputes, plagiarism, outdated publication, retraction of included studies, methodologic or data errors, and conflicts of interests. For 49 (13.0%), specific reasons were not provided. Of the retracted SRMAs, 41 (10.9%) were cited in CPGs; 19 (46.3%) of these SRMAs were retracted due to research integrity issues and 12 (29.3%), due to data errors or being outdated. For 10 (24.4%), specific reasons were not provided. Most retractions were due to manipulation of the publication or peer review process. The median time between publication and retraction of the SRMA used in CPGs was 12.0 (IQR, 3.5-25.0) months, and the median number of SRMA citations was 40 (IQR, 22-191). In the second stage, we included 36 CPGs of the 138 potential CPGs identified. Nine CPGs (25.0%) were retracted because of ethical reasons and 22 (61.1%) for nonethical reasons; the rest had no available information. The most common ethical reasons were plagiarism, authorship or intellectual property disputes, lack of disclosure of conflicts of interest, and discrepancies between the content and the cited evidence. Among the 22 CPGs retracted for nonethical reasons, 11 were due to dual publication or incorrect citations and 9 were due

to outdated recommendations. The median publication to retraction time was 10 (range, 3-96) months. All of these CPGs were cited after their retraction date, and in all cases, the citations were used to support the background of the research studies.

Conclusions Retracted SRMAs have been informing CPGs, which provide recommendations in practice and policy. The retraction of CPGs has been neglected. The RW database should be revised according to the specificities of CPGs. The most concerning reasons are ethical. Retracted CPGs continue to be cited after their retraction, mainly to inform the background sections of articles.

[1]Department of Pediatrics, University of Antioquia, Medellín, Colombia, ivan.florez@udea.edu.co; [2]School of Rehabilitation Science, McMaster University, Hamilton, Ontario, Canada; [3]Universidad del Norte, Barranquilla, Colombia; [4]Universidad Metropolitana, Barranquilla, Colombia; [5]EPICLINICA SAS, Barranquilla, Colombia.

Conflict of Interest Disclosures None reported.

Social Media

In-person

Video and Social Media Performance at a Surgical Journal With Video Journal Clubs

Caden Seraphine,[1] Abigail Chambers,[1] Susan Galandiuk[2]

Objective In recent years, *Diseases of the Colon & Rectum* (*DCR*), a single society-owned specialty journal, sought to grow its online presence. The journal launched a YouTube channel in 2018, featuring videos of monthly English-language journal clubs (JCs) discussing topical publications beginning in 2020, followed by article discussions in Spanish (2022-2023), Japanese (2024), and Korean (2024). The journal maintained its Facebook and X (formerly Twitter) presence and started LinkedIn and Instagram accounts in 2023. Articles were given a nonpaywalled status for 1 month, and upcoming discussions with article hyperlinks were announced 4 times per month via social media. Articles were accessible by hyperlinks, and article citations were displayed in JC discussant slides. We sought to explore how these activities impacted measures of journal interaction and global reach.[1,2]

Design We performed a 7-year (2018-2024) retrospective cohort study of JC video-based educational content provided by *DCR* with respect to journal website and social media access following STROBE guidelines.[3] We examined efforts toward global engagement, including addition of journal-specific non–English-language (2 Spanish, 1 Korean, and 1 Japanese) discussions. Analyses of 4 popular social media platforms' native analytics reporting and website engagement data examined viewer demographics (age, language, viewing country, and traffic source) and journal web-based interaction (number of views, number of unique visitors, and country of access) by month and year, trends in individual JC videos, and overall page engagement. We compared the median number of citations of JC articles (cumulative Web of Science citations) and their Altmetric scores (indicating article attention received) with mean citations of *DCR* articles per year.

Results Only 1 JC featured videos of surgical procedures; all others discussed original articles. The growth in YouTube engagement since the journal's YouTube channel launched in 2018 grew exponentially with the addition of monthly JCs in 2020. LinkedIn and Instagram were implemented in 2023, and cumulative traffic to the journal's website from social media platforms tripled within 1 year, from 15,414 clicks in 2023 to 56,609 in 2024 (Facebook, X, Instagram, LinkedIn, and YouTube). Throughout the period, individuals aged 25 to 34 years led YouTube viewership. International engagement increased with *DCR*'s implementation of non-English JC discussions in 2022. Publications discussed during JCs consistently had higher Web of Science citations compared with the mean citations in the journal that year (**Table 25-0959**).

Conclusions Younger audiences engage beyond traditional platforms of scholarly publishing in ways that hold inherent educational value. Journal Impact Factors exclude this reach and overlook the impact of web-based peer-reviewed educational content. Inclusion of social media and video-based learning tools, as well as efforts toward global reach, contribute to journal impact.

References

1. Narayan RR, Fleming AM, Gunder M, et al. Reflections from the *Annals of Surgical Oncology* social media committee: the impact of promoting surgical science online. *Ann Surg Oncol*. 2025;32(2):656-664. doi:10.1245/s10434-024-16420-4

2. Özkent Y. Social media usage to share information in communication journals: an analysis of social media activity and article citations. *PLoS One*. 2022;17(2):e0263725. doi:10.1371/journal.pone.0263725

3. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-349. doi:10.1016/j.jclinepi.2007.11.008

[1]Price Institute of Surgical Research, University of Louisville, Louisville, KY, US; [2]Division of Colorectal Surgery, Department of Surgery, University of Louisville, Louisville, KY, US, sogala01@louisville.edu.

Conflict of Interest Disclosures Susan Galandiuk receives a stipend for services as editor for the journal *Diseases of the Colon & Rectum* from the American Society of Colon and Rectal Surgeons. No other disclosures were reported.

Acknowledgments Margaret Abby (Managing Editor, *Diseases of the Colon & Rectum*; Research Grants Coordinator, Department of Surgery, University of Louisville, Louisville, KY), helped with obtaining data on articles discussed during journal clubs and their Web of Science citations.

Table 25-0959. Trends in Journal and Web-Based Video Content Engagement

Year (n)	Follower growth increase on all social media platforms	YouTube channel data			Web of Science citations		Altmetric score for English-language JC articles, median (IQR)
		Total YouTube channel views, No.	Countries represented in viewership, No.	Views from YouTube searches, No.[a]	Annual journal article citations, mean	English-language JC article citations 2020-2024, median (IQR)	
2020(n = 6)[b]	3385	11,712	11	4875	4.8	33.5 (20.0-42.3)	46.0 (22.3-54.8)
2021 (n = 20)	3985	131,762	45	84,950	4.7	16.0 (7.8-24.8)	17.5 (12.0-29.5)
2022 (n = 18)[c]	3703	166,390	55	87,500	4.1	17.0 (12.0-36.0)	25.0 (15.5-32.5)
2023 (n = 14)[c]	3614[d]	235,615	81	109,434	3.3	6.5 (4.0-10.5)	15.5 (13.0-17.8)
2024 (n = 14)[c]	8347	299,990	83	170,249	N/A	4.0 (2.0-27.8)	16.0 (4.0-24.5)[e]

Abbreviation: JC, journal club.

[a]Refers to views resulting from searches on the YouTube platform as reported by YouTube Studio Analytics.

[b]Expert discussion-based English-language journal club added October 2020.

[c]Non–English-language journal clubs implemented (Spanish, 2022 [Latin American] and 2023 [Europe]; Japanese, 2024; and Korean, 2024)

[d]Additional social media platforms instituted (LinkedIn, Instagram).

[e]Website posting of Altmetric scores discontinued on journal website in 2024.

Virtual

Altmetric Footprint of Retracted and Corrected Publications: The Role of Misinformation and Disinformation

Ashrah Maleki,[1] Niina Sormanen,[1] Kim Holmberg[1]

Objective While the alternative metric (altmetric) visibility of retracted publications has been studied,[1] less is known about how media and social media engage with corrections issued after retractions. This study examines how different types of retracted and corrected scientific publications engage scholarly and public audiences across altmetric platforms, focusing on how the reasons for retraction influence visibility. We distinguish between misinformation (unintentional error), disinformation (intentional deception), and disputed cases and assess their relative representation in altmetric activity.

Design This cross-sectional study analyzes altmetric data of retracted and corrected publications from multiple sources. Data were extracted in November and December 2024 from Retraction Watch, Scopus, and Altmetric.com (and updated in June 2025). The dataset comprised (1) 34,191 DOIs of original articles (later retracted), (2) 31,935 DOIs of retraction notices, (3) 1276 DOIs of correction notices, and (4) corrected versions (ie, retracted articles followed by a correction). To ensure clarity in altmetric attribution, we excluded records in which a single DOI was assigned to both the original article and its corresponding retraction or correction notice. Retraction reasons were coded using a 6-zone classification framework (based on Sormanen et al[2]) that distinguishes both intent and (nonscholarly) consequence: critical disinformation (eg, data falsification or fabrication), critical misinformation (eg, unreliable analyses or results), inconsequential disinformation (eg, duplication or plagiarism), and inconsequential misinformation (eg, miscommunication or lack of approvals), along with disputed (eg, expressions of concern) and ambiguous cases. Each publication was counted using a severity-exclusive method: when multiple reasons were present, the record was assigned to the most severe applicable category (**Table 25-1043**). Platform-specific mention counts were analyzed across 16 altmetric platforms, including news media, blogs, X/Twitter, Mendeley, Dimensions, and policy documents. Engagement levels were compared across document types, and how these varied by reason category were examined.

Results Among the 37,451 records, original articles received the most absolute attention: 41% had an Altmetric Attention Score (AAS) greater than zero, 43% were saved in Mendeley, and 28% appeared on X. Retraction notices were less visible (32% AAS > 0; 26% Mendeley). This means that retracted studies, often based on unreliable or deceptive findings, tend to receive more attention than the notices meant to correct them. Correction-related records showed higher proportional attention (51% of notices and 70% of retractions had an AAS greater than 0) but remained rare and largely absent from news (17% and 8%, respectively) and policy documents (6% and 1%, respectively). Among retraction notices, the most frequent reasons were critical disinformation (27%), inconsequential disinformation (27%), and critical misinformation (21%), yet these categories each saw only 8% to 10% with an AAS greater than 0.

Conclusions Retractions receive more attention than corrections, even in serious cases. Corrections remain mostly invisible in public-facing media, underscoring the need for better metadata and targeted communication strategies.

References

1. Serghiou S, Marton RM, Ioannidis JP. Media and social media attention to retracted articles according to Altmetric. *PLoS One*. 2021;16(5):e0248625. doi:10.1371.journal.pone.0248625

2. Sormanen N, Holmberg K, Maleki A. Characterizing retraction reasons: types of mis- and disinformation in the scientific publications. Research in progress.

[1]Department of Social Research, University of Turku, Turku, Finland, ashraf.maleki@utu.fi.

Conflict of Interest None reported.

Table 25-1043. Platform-Specific Engagement Across 4 Document Types (Retracted Articles, Retraction Notices, Correction Notices, and Corrected Versions) Based on Altmetric and Scholarly Metrics

Metric	No. (%)				Severity-based retraction reason category, No. (%) (n = 31,935)[a]				
	Retracted article	Retraction notice	Correction notice	Corrected retraction[b]	1. Critical disinformation	2. Critical misinformation	3. Inconsequential disinformation	4. Inconsequential misinformation	5. Disputes
AAS >0	14,050 (41)	10,184 (32)	651 (51)	888 (70)	2528 (8)	3226 (10)	2490 (8)	415 (1)	1241 (4)
Mendeley readers	14,749 (43)	8198 (26)	563 (44)	921 (72)	1889 (6)	2770 (9)	2026 (6)	326 (1)	939 (3)
Dimensions citations	14,213 (42)	4206 (13)	420 (33)	911 (71)	953 (3)	1600 (5)	886 (3)	142 (0.4)	406 (1)
X	9521 (28)	6582 (21)	328 (26)	530 (42)	1622 (5)	2139 (7)	1418 (4)	275 (1)	952 (3)
Blog	4801 (14)	4355 (14)	498 (39)	440 (34)	1201 (4)	1500 (5)	1000 (3)	153 (0.5)	426 (1)
News	2115 (6)	1316 (4)	218 (17)	98 (8)	326 (1)	512 (2)	222 (1)	65 (0.2)	169 (1)
Patent	1604 (5)	134 (0.4)	199 (16)	16 (1)	42 (0.1)	60 (0.2)	20 (0.1)	5 (0.02)	6 (0.02)
Facebook	1590 (5)	454 (1)	184 (14)	48 (4)	106 (0.3)	189 (1)	84 (0.3)	18 (0.1)	38 (0.1)
Peer review	1465 (4)	91 (0.3)	248 (19)	12 (1)	28 (0.1)	32 (0.1)	26 (0.1)	2 (0.01)	6 (0.01)
Wikipedia	978 (3)	1369 (4)	121 (9)	88 (7)	212 (1)	410 (1)	492 (2)	61 (0.2)	131 (0.4)
Policy	742 (2)	110 (0.3)	82 (6)	7 (1)	21 (0.1)	32 (0.1)	26 (0.1)	1 (0.00)	6 (0.02)
Reddit	339 (1)	105 (0.3)	40 (3)	9 (1)	14 (0.04)	51 (0.2)	7 (0.02)	6 (0.02)	27 (0.1)
Video	319 (1)	54 (0.2)	46 (4)	5 (0.4)	13 (0.04)	26 (0.1)	4 (0.01)	1 (0.003)	7 (0.02)
Clinical guidelines	487 (1)	39 (0.1)	29 (2)	2 (0.2)	2 (0.01)	11 (0.03)	3 (0.01)	3 (0.01)	3 (0.01)
F1000	351 (1)	6 (0.02)	69 (5)	3 (0.2)	1 (0.003)	3 (0.01)	2 (0.01)	NA	NA
Bluesky	161 (0.5)	66 (0.2)	26 (2)	6 (0.5)	13 (0.04)	31 (0.1)	4 (0.01)	8 (0.03)	10 (0.03)
Q&A[c]	53 (0.2)	9 (0.03)	13 (1)	3 (0.2)	1 (0.003)	6 (0.02)	NA	NA	2 (0.01)
Total	34,191 (100)	31,935 (100)	1276 (100)	1276 (100)	2640 (8)	3473 (11)	2672 (8)	446 (1)	1284 (4)

Abbreviations: AAS, Altmetric Attention Score; NA, not applicable.

[a]The severity-based retraction reason categories were classified exclusively as retraction notices in the Retraction Watch dataset.

[b]Corrected retraction: retracted article followed by a correction.

[c]Refer to mentions in questions and answers (eg, StackExchange).

Funding This research is supported by 2 projects: the European Media and Information Fund, focusing on unreliable science and the role of media regulation, and the Research Council of Finland, examining the impact of scientific misinformation.

Acknowledgment We sincerely appreciate Professor Mike Thelwall (University of Sheffield) for his support in acquiring and sharing the Retraction Watch dataset.

SPONSORS

GOLD

Wiley

wiley.com

Wiley champions those who see knowledge as a force for good. A trusted leader in research and learning, our pioneering solutions and services are paving the way for knowledge seekers as they work to solve the world's most important challenges. Around the globe, we break down barriers for innovators, empowering them to publish and advance discoveries in their fields, evolve their workforces, and shape minds through teaching and learning. Together, we are unlocking the creation and curation of knowledge for all, transforming today's biggest obstacles into tomorrow's brightest opportunities.

SILVER

ACS Publications

pubs.acs.org

ACS Publications' commitment to publishing high-quality research continues to attract impactful publications from top authors around the globe. A division of the American Chemical Society, ACS Publications supports researchers through journals, eBooks, scientific programs, and the news magazine Chemical & Engineering News. As a nonprofit scholarly publisher, ACS Publications offers, trusted peer review adjudicated by 1000 editors around the globe; a home for every type of research within a portfolio of more than 85 journals, including 12 "gold" completely open access journals; rapid publication within two weeks of acceptance; and broad global exposure with researchers at more than 5,000 institutions in 99 countries accessing ACS Publications. ACS looks forward to celebrating its 150th anniversary in 2026.

IEEE

ieee.org

IEEE is the world's largest technical professional organization and a public charity dedicated to advancing technology for the benefit of humanity. With our members, we inspire a global community through highly-cited publications, technology standards, conferences, and educational activities. IEEE provides authoritative thought leadership and creates an environment where professionals can collaborate on world-changing technologies—from computing and sustainable energy systems, to communications, healthcare, robotics, and more. For over 140 years, IEEE has been an essential resource for the global technical community as together we nurture, develop, and advance the building of technologies that improve lives.

New England Journal of Medicine

nejm.org

The *New England Journal of Medicine* is the world's leading general medical journal and website. Continuously published for over 200 years, NEJM publishes peer-reviewed research and interactive clinical content for physicians, educators and the global medical community. NEJM is a publication of NEJM Group, a division of the Massachusetts Medical Society, a nonprofit organization.

Wolters Kluwer Health Medical Research

wolterskluwer.com

Wolters Kluwer provides trusted clinical technology and evidence-based solutions that engage clinicians, patients, researchers, and the next generation of healthcare providers. Lippincott® Journals are at the forefront of the publishing innovation in digital channels, transforming how research is produced, distributed, accessed, and consumed for societies, authors, readers, and advertisers.

BRONZE

ACP/*Annals of Internal Medicine*

acpjournals.org

Annals of Internal Medicine, known for excellence in peer review, is a publication of the American College of Physicians. Annals promotes excellence, advances standards in research methods, and contributes to improving health worldwide by publishing research, reviews, guidelines, policy papers, and commentaries relevant to internal medicine and its subspecialties.

American Heart Association

heart.org

The American Heart Association (AHA) is the nation's oldest and largest voluntary organization dedicated to fighting heart disease and stroke. A shared focus on cardiovascular health unites our more than 35 million volunteers, supporters, and employees. The AHA invests billions in cardiovascular and cerebrovascular disease research.

American Society of Hematology

hematology.org

The American Society of Hematology (ASH) is the world's largest professional society of clinicians and scientists who are dedicated to conquering blood diseases. Since 1958, the Society has led the development of hematology as a discipline by promoting research, patient care, education, training, and advocacy in hematology.

Elsevier

elsevier.com

Elsevier, the global leader in advanced information and decision support for science leads the way towards systematic research on peer-review. Using Peer Review Workbench, researchers can access Elsevier journals manuscript metadata to perform peer review process analysis at scale, leveraging the powerful Databricks platform: all at no cost.

MPS/Highwire

highwirepress.com

HighWire is an industry leading global provider of digital publishing tools and platform solutions across all aspects of the publishing life cycle including content management and hosting, access and identity management, content authoring and editing, manuscript submission and tracking, e-commerce, and analytics.

Silverchair

silverchair.com

Silverchair is the leading independent platform partner for scholarly and professional publishers, serving our growing community through flexible technology and unparalleled services. Our teams build, maintain, and innovate platforms across the publishing lifecycle—from idea to impact. Our products facilitate submission, peer review, hosting, dissemination, and impact measurement, enabling researchers and professionals to maximize their contributions to our world.

EXHIBITORS

Aries Systems Corporation

ariessys.com

Aries Systems transforms and revolutionizes the delivery of high-value content to the world. We are committed to providing highly customizable, flexible, and innovative workflow solutions designed to help enhance the discovery and dissemination of human knowledge. Publish faster, publisher smarter, with Aries Systems.

BMJ Group

bmjgroup.com

BMJ Group is a global provider of medical information to clinicians, educators, policymakers and the medical research community. As publisher of The BMJ and 65+ further specialist journals, we are committed to driving innovation and standards in the conduct and publication of medical research.

Cactus Communications

cactusglobal.com

Cactus Communications is a trusted partner for publishers and societies, providing AI-driven and editorial solutions that enhance research discoverability, streamline workflows, and improve author experience. With cutting-edge technology and deep domain expertise, we help publishers scale efficiently, drive engagement, and maximize the impact of academic content.

Clarivate

clarivate.com

Clarivate is a leading global provider of transformative intelligence. We offer enriched data, insights & analytics, workflow solutions and expert services in the areas of Academia & Government, Intellectual Property and Life Sciences & Healthcare. Working with the scientific and academic community, we empower institutions and libraries to drive research excellence and student outcomes by connecting trusted content, deep expertise and responsible innovation. Clarivate is home to leading research solutions, including ProQuest, Web of Science, Pivot-RP, Esploro and InCites Benchmarking & Analytics.

Council of Science Editors

councilscienceeditors.org

The Council of Science Editors (CSE) is an international membership organization for editorial professionals publishing in the sciences. Our purpose is to serve our members in the scientific, scientific publishing, and information science communities by fostering networking, education, discussion, and exchange.

Dragonfly Editorial

dragonflyeditorial.com

Dragonfly is a B2B content agency, here to help you with writing, copy editing, graphic design, and content strategy. We're an extension of your team, easing your workload and reducing your stress. We create clear, compelling content that meets your business needs and helps you shine.

Hum

hum.works

Hum helps publishers create knowledge from content. Alchemist, the AI engine at the heart of Hum, doesn't just analyze data—it understands context, predicts trends, and creates intelligent connections between content and readers. From helping editors identify breakthrough research to delivering personalized content experiences, we're transforming how publishers operate in the digital age. Together with leading publishers worldwide, we're pioneering the intelligent publishing ecosystem of tomorrow.

JAMA Network

jamanetwork.com

Building on a tradition of editorial excellence, the JAMA Network brings *JAMA* together with *JAMA Network Open*, 11 JAMA Network specialty journals, and our new JAMA+ AI channel to offer enhanced access to research, reviews, and opinion shaping the future of medicine. Through a variety of publication and access options, innovative tools, and multimedia, the JAMA Network provides information and insights that matter most to medical research and practice. The JAMA Network journals are leaders in reach and impact, with more than 195 million article views per year.

KnowledgeWorks Global Ltd.

kwglobal.com

Improve your journal with exceptional editorial office and peer review support. KGL's people-centered, technology-enabled global team delivers outstanding quality and efficiency, now including the experts at Origin Editorial. With decades of experience supporting hundreds of journals, we bring leading-edge knowledge of industry best practices and a deep commitment to ensure your journal delivers the best possible service to your editors, authors, and reviewers.

Kriyadocs

kriyadocs.com

Kriyadocs is an ecosystem for scholarly publishers that streamlines end-to-end workflows from submission through peer review and production, delivering publication-ready content and metadata. Created by Exeter Premedia Services, Kriyadocs is committed to providing an exceptional experience for researchers and content publishers. The Kriyadocs publishing ecosystem offers a suite of technology solutions and services driven by our ethos of fostering agility in publishing and co-creating with the scholarly publishing community. Leveraging our wide range of solutions, we strive to advance scholarly publishing by promoting research integrity, detecting malpractice, and helping publishers deliver a great researcher experience.

MPS/Highwire

highwirepress.com

HighWire is an industry leading global provider of digital publishing tools and platform solutions across all aspects of the publishing life cycle including content management and hosting, access and identity management, content authoring and editing, manuscript submission and tracking, e-commerce, and analytics.

Newgen KnowledgeWorks

newgen.co

Our peer review service helps journals find the right reviewers quickly and easily. Using smart, data-driven tools, we match each manuscript with reviewers who have the right expertise—while avoiding recent collaborators to keep the process fair and unbiased. The platform automates the entire workflow: Sends personalized reviewer invitations; Manages follow-ups and retries; and Adjusts batch sizes for best results. This saves time for your editorial team and improves acceptance rates. You also get clear, easy-to-use dashboards that show how decisions are made—so you can trust the process and stay in control.

Proofig AI

proofig.com

Proofig AI is setting the global standard for scientific image integrity verification. Our diverse team brings together senior life science researchers, world-class computer vision experts, and seasoned SaaS innovators to transform how research is checked, verified, and published. Built for publishers, editors, researchers, and institutions, our AI-powered platform safeguards the authenticity of scientific images, helping to detect duplication, manipulation, plagiarism and AI-generated content before it reaches publication. Post publication detection is often too late and can lead to costly investigations, rejections and possible retractions. Our mission is simple yet critical: to protect the credibility of science**.** By delivering state-of-the-art, automated image analysis, we help ensure that every published figure meets the highest standards of accuracy, ethics, and trust.

Scholastica

scholasticahq.com

Scholastica is an academic publishing technology solutions provider with modern, easy-to-integrate software and services for peer review management, article production, and OA journal hosting. Our mission is to empower scholarly organizations to make quality research available more efficiently and affordably so they can further their missions. 1,300+ journals use Scholastica.

SciPinion

scipinion.com

SciPinion delivers high-quality peer reviews for organizations and publishers through the seamless integration of our expert community, proprietary technology and rigorous methodology. Our modified Delphi process ensures reproducible, reliable evaluations with unprecedented scale and speed. Our global pool of subject matter experts radically reduce editorial workload while enhancing scientific integrity.

Silverchair

silverchair.com

Silverchair is the leading independent platform partner for scholarly and professional publishers, serving our growing community through flexible technology and unparalleled services. Our teams build, maintain, and innovate platforms across the publishing lifecycle—from idea to impact. Our products facilitate submission, peer review, hosting, dissemination, and impact measurement, enabling researchers and professionals to maximize their contributions to our world.

Underline Science

underline.io

Underline Science is a leading provider of technology and services to fully support the delivery of virtual and hybrid conferences. We also originated the first video library of conference presentations and scholarly lectures, featuring the world's leading scientists and teachers. Each video is enriched to enhance the scholarly viewing experience.

University of Toronto Press

utppublishing.com

University of Toronto Press (UTP) is one of the largest university presses in North America, publishing landmark scholarship since 1901. UTP publishes over fifty leading journals in disciplines including microbiology, hepatology, physiotherapy, humanities, and social sciences, and each year releases over 200 new scholarly, course, and general interest books.

Wiley

wiley.com

Wiley champions those who see knowledge as a force for good. A trusted leader in research and learning, our pioneering solutions and services are paving the way for knowledge seekers as they work to solve the world's most important challenges. Around the globe, we break down barriers for innovators, empowering them to publish and advance discoveries in their fields, evolve their workforces, and shape minds through teaching and learning. Together, we are unlocking the creation and curation of knowledge for all, transforming today's biggest obstacles into tomorrow's brightest opportunities.

Tenth International Congress on Peer Review and Scientific Publication

Organizers and Planners

JAMA NETWORK

jamanetwork.com

Building on a tradition of editorial excellence, the JAMA Network brings JAMA together with JAMA Network Open, and 11 JAMA Network specialty journals to offer enhanced access to research, reviews, and opinion shaping the future of medicine. Through a variety of publication and access options, innovative tools, and multimedia, the JAMA Network provides information and insights that matter most to medical research and practice. The JAMA Network journals are leaders in reach and impact, with more than 135 million article views per year.

BMJ

bmj.com/company

Global medical publisher BMJ supports its vision of a healthier world by sharing knowledge, evidence, and expertise that improve health outcomes. Our medical journal portfolio includes our flagship, The BMJ, and 70+ open access and hybrid specialty titles in oncology, pain medicine, and more. Launched in 2022, BMJ Medicine is a new open access, multispecialty journal from The BMJ that aims to improve clinical practice, policy, and medical science. In addition, we are a co-founder of medRxiv and offer digital clinical decision support tools that help healthcare professionals improve the quality of healthcare delivery.

Meta-Research Innovation Center at Stanford (METRICS)

metrics.stanford.edu

The Meta-Research Innovation Center at Stanford (METRICS) is a research to action center focused on transforming research practices to improve the quality of scientific studies in biomedicine and beyond. METRICS fosters multi-disciplinary research collaborations to help produce solutions that increase the effectiveness and value of scientific investigation.

Thank you!

The organizers wish to thank all plenary session and poster session presenters, moderators, advisory board members, sponsors, exhibitors, and all participants, and the following individuals who contributed to the planning and support of this Congress: Jeni Reiling, Rosa Miranda, Caroline Sietmann, Kirby Snell, Joe Amft, KC Walsh, Jose Santa Jr, Sreeparna Bose, Craig McCaffrey, Rick Bell, Amber Reynolds, Sherry Flores, Morgan Osgood, Erin Kato, Andrew Given, Nicole Iwinski, Lori Ramos, Ted Grudzinski, Debra Camp, Jacob Kendall-Taylor, Anna Wietraszuk, Michaela Mark, Sylvia Orellana, Sara M. Billings, Kevin Brown, Nicole Fiorito, Timothy Gray, Bernadette Hromin, Emma Hyche, Kate Lander, Peter Olson, Juliet Orellana, Jen Phillis, Paul Ruich, Kristine Simmons, and Shannon Sparenga.

With Sponsorship From

GOLD

Wiley
wiley.com

SILVER

ACS Publications
pubs.acs.org

IEEE
ieee.org

New England Journal of Medicine
nejm.org

Wolters Kluwer Health Medical Research
wolterskluwer.com

BRONZE

ACP/*Annals of Internal Medicine*
acpjournals.org

American Heart Association
heart.org

American Society of Hematology
hematology.org

Elsevier
elsevier.com

MPS/Highwire
highwirepress.com

Silverchair
silverchair.com

JAMA Network

thebmj

METRICS
META-RESEARCH INNOVATION CENTER AT STANFORD