

## Title

Analysing Elsevier Journal Metadata with a New Specialized Workbench inside ICSR Lab

## Authors

Ramadurai Petchiappan<sup>1</sup>, Data Science Research Content (DSRC), Elsevier, Chennai, India

Kristy James<sup>2,4</sup>

Andrew Plume<sup>2,4</sup>

Efthymios Tsakonas<sup>3</sup>, Data Science Research Content (DSRC), Elsevier, Amsterdam, The Netherlands

Ana Marušić<sup>5</sup> (ORCID: <https://orcid.org/0000-0001-6272-0917>), Center for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia

Mario Malički<sup>6</sup> (ORCID ID: [0000-0003-0698-193](https://orcid.org/0000-0003-0698-193)), Meta-Research Innovation Center at Stanford (METRICS), Stanford University

Francisco Grimaldo<sup>7</sup> (ORCID ID: [0000-0002-1357-7170](https://orcid.org/0000-0002-1357-7170)), Department of Computer Science, University of Valencia, Spain

Bahar Mehmani<sup>3</sup> (ORCID <https://orcid.org/0000-0003-4038-4531>), STM Journals, Elsevier, Amsterdam, The Netherlands

<sup>1</sup> International Tech Park, Taramani Road, Taramani, CHENNAI, 600113, India

<sup>2</sup> International Center for the Study of Research, Elsevier, Radarweg 29, Amsterdam, The Netherlands

<sup>3</sup> Elsevier, Radarweg 29, Amsterdam, The Netherlands

<sup>4</sup> Elsevier, The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, United Kingdom

<sup>5</sup> University of Split School of Medicine, Department of Research in Biomedicine and Health, Center for Evidence-based Medicine, Šoltanska 2, 21000 Split, Croatia

<sup>6</sup> Stanford University, Medical School Office Building, 1265 Welch Rd, Stanford, California 94305

<sup>7</sup> University of Valencia, Department of Computer Science, Av. de la Universitat s/n, 46100, Burjassot, Spain

## Abstract

In this white paper we introduce Elsevier's Peer Review Workbench which will be available via the computational platform ICSR Lab. The workbench offers a unique dataset to interested researchers who want to run research on journal evaluation and peer review processes. We describe its properties, advantages, and limitations as well as the process of proposal application. This is a living document and will be updated on a regular basis.

## Introduction

Peer review is the cornerstone of academic evaluation and scholarly communication: it can help improve publications and their credibility, ensure quality and integrity of research, further knowledge generation and exchange, and increase trust in control mechanisms set by the content providers (e.g. journals, societies, publishers, platforms, etc.).

The concept of peer review has changed throughout time [1,2]. Initially meant for the evaluation of manuscripts submitted to journals, it is now being used in a broader sense and at different stages of research cycles [3]. It happens at the early stages of research through evaluation of grant applications to validation and commenting processes on preprint services all the way to new research outputs such as data, protocol, and software.

Despite its importance, research on peer review is scarce and fragmented [1]. This also explains why management of scholarly journals mostly revolves around a set of informal, ground-up practices, rarely supported by evidence-based, systematic analysis. Current investigations are fragmented, with few connections and limited knowledge-sharing, as manifested by how sparsely these researchers cite each other's papers. The topic is under-studied partly because it is difficult to research. Access to data about the review process is difficult because of the confidentiality of the process and difficulties in anonymizing datasets. It often depends on personal connections with journals, and is generally limited to such a small number of titles making generalizations difficult. Few dedicated grants are available. Yet greater transparency and studies could help determine which models and practices of peer review are most effective in reaching their desired goals.

For decades, academics have called for study and research on peer review [4]. Elsevier is now leading the way towards systematic research on peer-review by launching the Peer Review Workbench (PRW) in ICSR Lab. PRW is an environment for providing responsible access to academics to the enhanced and enriched Elsevier journal and manuscript metadata inside ICSR Lab (<https://lab.icsr.net/about.html>) dedicated to evidence-based research on peer review.

Researchers and academics can apply to access metadata for manuscripts in Elsevier journals to run systematic analysis on peer review processes in different disciplines at scale.

## Data

The peer review workbench data includes metadata from Elsevier proprietary journal manuscripts processed 2018-May 2021 (will be updated annually). Though for anonymization purposes only a subset of these fields will be available for any one project, the underlying dataset includes:

<b>Serial. No</b>	<b>Metadata</b>
<b>Submission data</b>	
1	Journal code
2	Manuscript code
3	Article type
4	Submission date
5	Status of submission
6	Decision
7	Decision date
8	Keywords
9	Abstract (Only for accepted manuscripts upon further agreement.)
10	Number of co-authors
11	Authors' and Co-Authors' gender*
<b>Review data</b>	
12	Reviewers' comments to author(s) (Only for accepted manuscripts upon further agreement.)
13	Reviewers' invite-accept date
14	Review completion date
15	Review status
16	Decision recommendation
17	Reviewers' gender*
18	Reviewers' geographical location
<b>Editorial decision data</b>	
19	Decision date
20	Editor's gender*
21	Author position
22	Board member
23	Editor's Institution
24	Editor's City
25	Editor's State
26	Editor's Country

27	Editor's Classification
<b>Journal metadata</b>	
28	Subject area classification
29	Impact factor quartile

\*Gender assignment metadata was derived using an AI-driven methodology that is appropriate for bibliometrics or other large-scale analyses because such studies focus on trends at scale [6]. However, the methodology does not guarantee that an individual's gender can be identified with flawless precision; thus, the gender metadata cannot be used for an individual level or small group analyses as an alternative to self-reported data.

It is worth clarifying that no author, reviewer, and editor identity will be available in the dataset. Though they are available at the stage of internal processing, they are removed before the data is made available in the Peer Review Workbench.

The Peer Review Workbench data does not include personal identifiers, nor does ICSR Lab host article full text. As for the reviewer-comment-to-author text, for data privacy reasons, we do not include the text in the dataset in ICSR Lab. We foresee that studies that require text analysis will follow a procedure in which researchers submit their analysis scripts to be run by Elsevier staff in our internal data analysis platform on researchers' behalf. This will allow for a study design that includes variables such as e.g. counts of certain keywords or the length or sentiment of comments but prevents these texts being read at an individual level. Handling proposals for studies on these texts require a separate mechanism and is not included in our initial call for proposals for the PRW.

To ensure that the peer review data remains anonymous and can only be viewed at an aggregate level, each project will receive access to a pre-processed subset of this dataset with the fields required to answer their research question, as determined in conjunction with the Peer Review Workbench advisory board. Identifiers such as journal ID or manuscript ID will be obscured or omitted. If there are any numerical variables added that could be used to identify an individual submission (such as citations accrued by the article) this will be masked by using quartile or decile information instead of the real number. Through pre-processing the subset, we try to minimise the risk of individual persons' identification. In addition, PRW projects cannot access other datasets apart from those directly required for their proposed work, in order to further prevent risk of de-anonymization.

Please note that we do not encourage studies solely aiming at measuring journal publication speed or acceptance rates as these are often available on journal home pages (e.g. <https://www.journals.elsevier.com/current-applied-physics> and <https://www.journals.elsevier.com/water-research>).

### **Breakdown of dataset by field**

While we are still working on the a more complete summary and breakdown of the dataset and will update this section in upcoming versions, we report on our plan of reporting and items we feel confident in this current version. For example, counts of the multidisciplinary items needs to be validated. Table 1 summarizes our current report and provides an overview of what to expect in the next version of this document.

Table 1. Summary data of the Peer Review Workbench Dataset

Variable	Health Sciences	Life Sciences	Physical Sciences	Social Sciences	Multidisciplinary	Total
No. of journals	288	244	619	192		1600
Journals without IF (no, %)						
No. of manuscripts	588191	519271	2656521	414576		4300640
Time from submission to first review decision (Md, IQR) in days	(24,34)	(23,31)	(29,40)	(47,60)		(29,41)
Authors per manuscript(Md, IQR)	(5,5)	(5,4)	(4,3)	(3,2)		(4,3)
Gender ratio of authors per manuscript						
No of first round reviewers per manuscript(Md, IQR)	(0,4)	(0,4)	(0,5)	(0,3)		(0,4)
Gender ratio of first round reviewers per manuscript						
No of reviewers invited in first round (Md, IQR)	(0,5)	(0,6)	(0,6)	(0,4)		(0,5)
No (%) of manuscripts with more than 1 review round	12%	17%	17%	13%		15%
Inter Rater Review agreement for round 1						
Time from invite to acceptance of reviewers (Md, IQR) in days	(9,16)	(8,15)	(14,23)	(21,35)		(11,19)
Unique editors						

gender (No, %)						
Manuscript Acceptance rate	22%	24%	22%	13%		21%

## Gender inference using Namsor API

Among the several approaches we experimented for gender inference, we found that most promising results were obtained using Namsor [7]. Among its other features, Namsor's API can help determine if a name is more likely to be perceived as male or female; and this estimation is accompanied with a confidence score. More details around this API can be found at [7].

## Potential studies

Rejection studies:

1. What happens to rejected manuscripts? Are they subsequently published elsewhere, or not at all (at least, in some recognisable form)? If they are published elsewhere, are they published in a more generalist or specialist title? Does that journal have a higher or lower ranking in terms of perceived prestige or typical research metrics? Are rejected manuscripts associated with certain author career stages, gender, institutional or national background?
2. The relationship between reviewers' recommendation and the editorial decision .
3. Are specific keyword combinations more likely to lead to a rejected paper?

Review effect on manuscripts:

1. Do (additional) rounds of review improve manuscripts? Are those accepted after a single round different in any way from those that succeed after several rounds?
2. What is the median or average number of review invites per manuscript?

Under-served topics:

1. What topics are typically – or increasingly – received as manuscripts but rejected from publication?

Inter- and multi-disciplinary and peer review:

1. Are highly inter-disciplinary manuscripts more or less likely to be accepted for publication? Do they require more rounds of review, or more reviewers, than purely disciplinary papers?

Inclusion, Equity, Diversity, and accessibility in the journal peer review process:

1. Are there any differences between reviewer performance from different countries, career levels, gender, etc. (see for example our recent study: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3912607](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3912607))
2. Are there any differences in gender or geographical location of authors more associated with rejections/acceptance?

3. The role of editor, authors, and reviewers' gender in peer review (see for example our study: <https://doi.org/10.1126/sciadv.abd0299>)

## How to apply

Researchers can apply to work with the Peer Review Workbench via the application process to ICSR Lab, which is an online form that can be accessed from <https://www.elsevier.com/icsr/icsrlab/how-to-apply>.

This application includes a form that collects contact information from the researchers who will require access, as well as a detailed abstract outlining the proposed research project and methodology.

For the PRW dataset we also ask additional details: Study type and protocol or reporting guideline to be followed; sample size calculations; dataset attributes required for the planned analysis; statistical analysis plan; statement of competing interests; funding information; code availability statement; and where relevant, we recommend including a statement on inclusivity and diversity and how this will be addressed in the project. We also highly encourage pre-registration of these proposals.

Following this application, we submit all applications to a technical review (to assess whether the platform has the capacity to analyse the data in the way described) and peer review – in the case of access requested of the PRW. This is conducted by a board of experts convened in 2022, consisting of six academics with experience in peer review research. The board will evaluate applications on a quarterly basis. We highly encourage multi-analyst studies, or group applications with members from diverse geographical regions. In cases where multiple teams apply to analyse the same research question, we will aim to introduce the teams to each other and propose they consider working together or in parallel (i.e. replication study). Researchers or teams whose proposals are accepted will then be given access to relevant data.

## Application timeline in 2022-2023

Call for proposals opens on 8 September 2022 with the deadline being 10 January 2023. Proposal evaluation will take approximately 4 weeks and by March 2023 the applicants will receive the decision letter. The advisory board will encourage collaboration among the applicants in case they see fit and request them to re-submit their proposals based on the proposed collaborations within one month.

PRW access will be provided to the accepted and finalized proposals as of 2 April 2023.

The proposal call cycle will resume on September 2023 with a the similar format expressed above.

## Application evaluation criteria

The criteria against which applications are evaluated are:

**Person-based criteria:** All applicants must have an academic affiliation; (in the case of PRW) teams must have at least one member who is a domain expert in scientometrics/peer review studies, as demonstrated through previous work; at least one member must have robust coding experience and willingness to learn Python/SQL and [PySpark](#) in order to query this data on our computational

platform; and the proposed project timeline must be judged appropriate to our review, acceptance and onboarding timelines (for which reason we do not recommend applications to complete Graduate or Master's thesis work with the PRW). We highly encourage multi-analyst approaches or replication teams.

**Research question criteria:** The research objective must be novel, or if replicating previous work discuss this appropriately; it must be an academic question relating sufficiently to the ICSR research themes<sup>1</sup> or advance the field of peer review studies; it must be sufficiently answerable with the datasets available (taking into account the coverage of the datasets and their potential coverage bias), or detail other supporting methods that will complement the data analysis conducted within the PRW. Note that if projects apply in the same call with similar questions, we will offer whether they would like to share their contact details in order to align their efforts and collaborate on the question.

## Acknowledgement

The preparation of this dataset is a result of a great team work by many people from different departments including Elsevier legal and data privacy Alex Chan, IJbrand Jan Aalbersberg, and Jessica Alexander; ICSR team, Alick Bird, Silvia Dobre, Rachel Herbert, and Peter Darroch; Aries and Elsevier submission systems team, Natarajan Kaliyamoorthy, Lauren Oppenheim, Rich Remington, Navid Bazari, Senthil Annamalai Ramakrishnan, and Ramsundhar Baskaravelu; Elsevier data science research content team Pascal Coupet, and Georgios Tsatsaronis; Elsevier Research Collaboration Unit, Noelle Gracy and Pragya Singh.

## References

- [1] Fyfe A, Gielas A. Introduction: Editorship and the editing of scientific journals, 1750–1950. *Centaurus*. 2020;62(1):5-20. <https://doi.org/https://doi.org/10.1111/1600-0498.12290>
- [2] Burnham JC. The evolution of editorial peer review. *Jama*. 1990;263(10):1323-9.
- [3] Tennant, J.P., Ross-Hellauer, T. The limitations to our understanding of peer review. *Res Integr Peer Rev* 5, 6 (2020). <https://doi.org/10.1186/s41073-020-00092-1>
- [4] Squazzoni, F., et. al. (2020). Unlock ways to share data on peer review. *Nature*, comment 25 Feb. 2020; <https://www.nature.com/articles/d41586-020-00500-y>
- [5] Grimaldo, F., Marušić, A., & Squazzoni, F. (2018). Fragments of peer review: A quantitative analysis of the literature (1969-2015). *PloS one*, 13(2), e0193148. <https://doi.org/10.1371/journal.pone.0193148>
- [6] Elsevier's report on Gender in Research: <https://www.elsevier.com/connect/gender-report>
- [7] <https://namsor.app/>

---

<sup>1</sup> These are (as of August 2022), impact of research, inclusivity, Open Science, research careers, research globalization, research practices and sustainability. See <https://www.elsevier.com/icsr/research-themes>.