**Guide to meta-analysis files for Smith et al. (*in review*)**

**Key notes:**

* Each review stage has its own .csv file, separated by (1) bias without interventions, (2) double-blind vs. single-blind, (3) Editor homophily, and (4) reviewer homophily. Demographics and author positions are pooled in a file. The R code separates demographics/positions prior to analysis. There are three exceptions:
  + Data broken down into individual countries are separated into their own review stage files. This is because each country has additional information. These files are designated as “[REVIEW STAGE] [DATASET] COUNTRY”
  + For journal nationalism/journal country, all stage data are combined by this demographic category into one file with binomial outcome data (“Journal nationalism problem binomial.csv”; pre-initial, initial, post-initial, final, and overall decisions) and one with non-binomial outcome data (“Journal nationalism problem nonbinomial.csv”; review scores, time of review, number reviewers, and number revisions). This is because the file has extra information pertaining to the journal location.
  + The number of submissions files have one file with all data, “Number submissions data problem.csv,” and a series of 6 files created to allow for Likelihood Ratio Tests. The authorship positions are divided into different files due to the data formatting differing from other review aspects. The six include
    - “Number submissions data FIRST CONTINENT problem.csv” that has complete observations for the first author’s affiliation’s continent
    - “Number submissions data CORR CONTINENT problem.csv” that has complete observations for the corresponding author’s affiliation’s continent
    - “Number submissions data LAST CONTINENT problem.csv” that has complete observations for the last author’s affiliation’s continent
    - “Number submissions data FIRST COUNTRY problem.csv” that has complete observations for the first author’s affiliation’s continent, language, and Human Development Index
    - “Number submissions data CORR COUNTRY problem.csv” that has complete observations for the corresponding author’s affiliation’s continent, language, and Human Development Index
    - “Number submissions data LAST COUNTRY problem.csv” that has complete observations for the last author’s affiliation’s continent, language, and Human Development Index
* Data examining bias without considering interventions is labelled “problem”
* Data examining impacts of possible solutions is labelled “solution”

**R code guide:**

* Code for analyses looking at bias in the absence of interventions are divided into review stages/aspects that are binomial (rejected, not rejected) and stages/aspects that are NOT binomial
  + “Problem-end analyses for binomial data.R”
    - Pre-initial, initial, post-initial review, and overall decisions
  + “Problem-end analyses for NON binomial data.R”
* Code for analyses examining efficacy of double-blind review is in a single file, “Solution-end analyses DOUBLE-BLIND.R”
* Code for analyses examining assessor homophily is in a single file, “Solution-end analyses HOMOPHILY.R”

**Table S2 from Online Supplementary Materials.** Definitions of review stages, authorship positions, and demographic classifications used in our quantitative meta-analysis.

|  |  |
| --- | --- |
| **Item** | **Description** |
| Assumed gender | In most (but not all) studies, gender was assumed using “gender-guessing” software or by manually assuming gender by going through first names (and sometimes using web searches for people). Generally, the studies included in our meta-analysis excluded manuscripts by authors for which the names were not above a certain probability of being “male” or “female.” Note that we are primarily examining assumed gender based on first names rather than actual gender.  For reviewer diversity studies, sometimes studies used the individual reviewer as the unit of replication and summarized scores for each reviewer/author interaction. In other studies, they used the manuscript as the unit of replication. In those cases, reviewer gender was classified as “1+ female” (1 or more reviewers were assigned female) and “all male” (all reviewers were assigned male). |
| Author position | Studies looked at a variety of authorship positions. We conducted analyses using first, corresponding, and last authors due their representation across datasets. We recategorized author positions to first, corresponding, and/or last in studies when the study indicated there was ~90% or higher overlap. |
| First = first listed author |
| Corresponding = listed as corresponding author |
| Last = last listed author |
| All = study assigned all authors of a manuscript to that demographic category. For these studies, we duplicated the rows for first, corresponding, and last author, given that those positions would all be assigned to that demographic. |
| First and/or corresponding = study assigned the paper to that demographic if there was at least one first or corresponding author that were assigned to that demographic category. We re-assigned these papers to first or corresponding if there was an ~90% or higher overlap with these categories. This occurred in Goldstone et al. (2020) that was reassigned into first and Zhang (2012) that was reassigned to both given that ~90% of papers had 100% authors from the demographic being studied (Chinese authors), with the remaining having first and/or corresponding. |
| Blinding format/peer-review model | Single-blind review = the reviewers know the identity of the authors, but the reviewer identities remain unknown to the authors. |
| Assumed single-blind review = journals did not specify a review model but the title page with author details was part of the main manuscript document |
| Double-blind review = both the reviewer and author identities remain unknown to each other. We separated double-blind review into (1) optional double-blind wherein the journal allows the authors to choose to remain anonymous or be known to reviewers and (2) required double-blind wherein the journals require authors to remain anonymous to reviewers. |
| Triple-blind review = the author's identities are unknown to the editors until after the initial decision and the author and reviewer identities remain unknown to each other throughout the review process. |
| Open review = mandatory identification of reviewers at any point during the review or publication process. |
| Other review model = the review models were not single-, double-, triple-blind, or open review. For example, some journals have a public commentary period on a preprint server. We included the Frontiers Media SA journals in “other” since reviewers that endorse the article are named on the published article but can opt out of being named if they withdraw or do not endorse a publication. Additionally, Frontiers Media SA journals use “collaborative peer review.” |
| Continent classifications | We classified continents following the CIA World Factbook except that we separated the Americas into Latin America and North America rather than South America and North America. In some cases, articles provided data by continent but not by country, in which case, we used what the study provided. |
| English as a primary language | We assigned English as a primary language using the CIA World Factbook following Burns and Fox (2017). That is, we assigned English as a primary language of a country if English was the first listed language or an official language. We separated English as a primary language data into comparisons for (1) countries within 0.10 Human Development Index points at the time of the study (similar development) and (2) all countries examined (any development index). Some studies only provided already aggregated English as a primary language, so in those cases, we used the values provided by the study and assigned them to each group as appropriate. |
| Final decision | If manuscripts are accepted or rejected after peer review. This stage only includes manuscripts that got past the initial decision stage. |
| Journal country or journal nationalism | A type of data looking at outcomes for authors depending on if they are from the country the journal is based in or from another country |
| Initial decision | First Editor decision to immediately reject manuscripts or send them to peer review. |
| Number of journals submitted to or “number of submissions” | The number of journals a single manuscript was submitted to prior to being accepted at one. These data all came from Fox and Paine (2019) wherein they sent a questionnaire survey to authors on published manuscripts tracing the history of those manuscripts prior to eventual publication. |
| Number of reviewers | Mean number of unique peer reviews received per article. |
| Number of revisions | Mean number of revisions per manuscript across the review process. |
| Overall decision | If a paper is accepted or rejected at any time during the peer review process. If it was immediately rejected (not sent into review) or rejected after 1 or more rounds of review, it was marked as rejected. If it is accepted at any point, it was marked as accepted. |
| Post-initial review | Decision to reject, revise, accept, etc. following the first round of peer review. |
| Prestige | Studies included in our meta-analysis classified prestige in varied ways.  Individual – This was classified by 1) H-index of the last author; and 2) being a member of the Society of Experimental Social Psychology (which requires nomination and a vote for admission)  Institutional – This was classified by 1) monetary value of research and training grants and contracts funded by the NIH; 2) Times Higher Education rankings; 3) the average of a scholarly productivity rating and institutional quality rating; and 4) 2012 Shanghai academic ranking of world universities for the life sciences. |
| Review scores | Review scores were converted into a semi-continuous scale from 0 (perfect acceptance) to 1 (perfect rejection). Some studies examined ordinal reviewer recommendations, which varied by study and journals included therein, but followed a format such as 1 = accept, 2 = minor revision, 3 = major revision, 4 = reject; 1 = accept, 2 = major or minor revision, 3 = reject; 1 = accept or minor revision, 2 = major revision, 3 = reject. Some studies included in our meta-analysis only provided review score data that had already been converted into weighted ordinal scores. That is, studies multiplied each decision’s score by the number of manuscripts in that category, then divided by the total number manuscripts. For example, if there were 4 accepted papers, 6 papers with a minor revision, 50 papers with a major revision, and 18 papers with a rejection, the score would be calculated as (4\*1 + 6\*2 + 50\*3 +18 \*4)/(4+6+50+18). Some studies only provided weighted ordinal scales and some studies only provided the number of recommendations per manuscript, so we calculated weighted ordinal scales in the latter case. Other studies provided reviewer assessment data on continuous scales such as a reviewer priority rating from 0 to 100. We reverse coded any scales that went from lower = worse to higher = better. We re-scaled all values to range from 0 to 1 by taking ([value] – [min possible value])/([max possible value] – [min possible value]). |
| Sample size for ordinal scale scores | Total papers included in the reviewer/editor scores. |
| Time of peer review | Mean time for peer reviewers to return reviews. |
| Total published | Total articles published by authors assigned to that demographic over the time period examined. This metric was used in before/after double-blind studies or studies examining the number of manuscripts published in similar paired single-blind/double-blind journals. These studies are summarized in the Supplementary Materials and were not used in our quantitative meta-analysis. |
| Pre-initial decision | Editor decision to send a manuscript to an Associate Editor, which was only presented by Fox et al. (2016) and Burns and Fox (2017). |
| Time to decision | Time in days from submission to first decision by the editor, provided that the paper was sent to review. |
| Time to desk rejection | Time in days from submission to desk rejection (immediate rejection without review) by the editor. |
| Time to post-review decision | Time in days from when the editor received the last peer review to when they made their post-initial review decision. |

Column names and descriptions for variables included in meta-analysis data files.

|  |  |
| --- | --- |
| **Column** | **Description** |
| % rejected | Percent rejected for that demographic for that stage for that study |
| Accept | Used in reviewer scores datasets. The number of manuscripts recommended for “acceptance” by reviewers. Studies were conducted on journals with a wide variety of actual possible categories/labels, so details on the actual labels from the original studies are in “Notes” |
| Assessor | Used in homophily dataset to indicate if the data pertain to editors or reviewers |
| Author country | Used in the journal nationalism dataset and the reviewer homophily dataset. Country where the respective author’s affiliation is located. The data available from studies on these areas were often provided in differing levels of details. Often, studies were conducted on focal countries (often the United Kingdom or United States), and data were coded as “[FOCAL COUNTRY]” or “[NOT FOCAL COUNTRY].” In the case of Opthof *et al.* (2002), data were also provided in the original study as “Match” and “Not match” across all countries. |
| Author match | If author location and journal location are the same = match, if not the same = not match. Used in place of “demographic.category” for analyses on journal nationalism and reviewer country homophily. |
| Author position | First – first author listed on manuscript. This is usually the person that led the work.  Corresponding – author designated as corresponding on manuscript.  Last – author listed last on manuscript. In many disciplines within the biological sciences, this is the person that is the lab head.  See Table S2 in the Supplementary Materials (reproduced above) for more information |
| Blinding | Indicates if double-blind review is optional (“Optional\_double”) or required (“Required\_double”) for journal(s) in that study. NA if not indicated by the study. |
| Blinding format | Peer review model of the journal(s) during the year(s) the data correspond to. |
| Category | Biological Sciences sub-category for journals included within that study:  Broad = covered multiple fields but publishes in the biological sciences like *Nature* (correspond to Tier 2 journals in our journal policy data)  Health sciences = medical journals  Life sciences = biological sciences outside of health sciences  We used the majority type of journal if more than one was used in that study. In the case of Squazzoni et al. (2021), we maintained the health sciences/life sciences breaks provided in their original dataset since the study covered >100,000 manuscripts and 79 journals and could be broken up easily. |
| Continent | Used as a column in data at the individual country level. Continent of the respective author’s affiliation’s country. Assigned using the same scheme reported elsewhere. |
| Continuous data | Used in nationalism dataset. Data corresponding to respective review stage (e.g., for number of revisions, this is the number of revisions; for number reviewers, this is the number of reviewers; for reviewer scores, this is the review score not rescaled to be between 0 and 1) |
| corr\_author\_country | Used in the number of submissions database. Country of the corresponding author’s affiliation on the manuscript. |
| corr\_author\_gender | Used in the number of submissions database. Assumed gender of the corresponding author. |
| Corr\_continent | Used in the number of submissions database. Continent matching the country of the corresponding author’s affiliation on the manuscript. |
| Corr\_HDI2012 | Used in the number of submissions database. Country Human Development Index matching the country of the corresponding author’s affiliation on the manuscript during the mean study year (2012). |
| Corr\_language | Used in the number of submissions database. Country primary language matching the country of the corresponding author’s affiliation on the manuscript. |
| Corr\_language\_similar\_HDI | Used in the number of submissions database. Country primary language matching the country of the corresponding author’s affiliation on the manuscript. This column only includes rows whose author’s country was within 0.10 HDI points of the highest HDI country in that dataset to control for HDI possibly confounding language effects. |
| Country | Country of the respective author’s affiliation’s location |
| Demographic | Demographic group that data pertain to (e.g., continent, gender).  Continent = continent of the author’s affiliation’s country.  ESL\_all = primary language for the country of the author’s institutional affiliation. This category includes authors from all countries, regardless of country Human Development Index (HDI).  ESL\_similarHDI = primary language for the country of the author’s institutional affiliation. This category only includes authors within 0.10 HDI points of the highest HDI country in that dataset to control for HDI possibly confounding language effects.  Gender = author’s assumed gender.  Prestige\_author = author prestige metric; see Table S2 in Supplementary Materials for details on how this was calculated for individual studies.  Prestige\_institution = institutional prestige metric for the institution listed in the author’s affiliation; see Table S2 in Supplementary Materials (reproduced above) for details on how this was calculated for individual studies. |
| Demographic category | Specific demographic (e.g., “Asia” if “Demographic” is “continent”). “Z\_” was used to assign the more privileged group as the reference group for analyses. “Z\_” may not appear in data lines that were just summarized due to low replication. |
| Editor demographic | Editor demographic category following same labelling scheme as author demographic category |
| English.fluency | Used in the number of submissions database. Self-identified English fluency of the corresponding author from the author questionnaire survey that the data correspond to. Authors could select “native speaker,” “very fluent,” “somewhat fluent,” “slightly fluent,” or “no knowledge of English.” |
| English.fluency.1 | Used in the number of submissions database. Binary version of “English.fluency” where “native speaker” was coded as “English” and all other categories were coded as “Not English.” |
| first\_author\_country | Used in the number of submissions database. Country of the first author’s affiliation on the manuscript. |
| first\_author\_gender | Used in the number of submissions database. Assumed gender of the first author. |
| First\_continent | Used in the number of submissions database. Continent matching the country of the first author’s affiliation on the manuscript. |
| First\_HDI2012 | Used in the number of submissions database. Country Human Development Index matching the country of the first author’s affiliation on the manuscript during the mean study year (2012). |
| First\_language | Used in the number of submissions database. Country primary language matching the country of the first author’s affiliation on the manuscript. |
| First\_language\_similar\_HDI | Used in the number of submissions database. Country primary language matching the country of the first author’s affiliation on the manuscript. This column only includes rows whose author’s country was within 0.10 HDI points of the highest HDI country in that dataset to control for HDI possibly confounding language effects. |
| Gender | Used in the number of submissions database. Self-identified gender of the corresponding author from the author questionnaire survey that the data correspond to. |
| HDI | Human Development Index (HDI; only used in the individual country-level datasets) of the country of the respective author’s institutional affiliation. This corresponds to the HDI during the mean study year. However, the HDI was started in 1990, so any studies conducted prior to that use the 1990 HDI. In some cases, countries do not have HDI data for the year corresponding to the study (e.g., occupation by another country), so we used the first available year in those cases. Full HDI data can be downloaded from the United Nations Development Programme: https://hdr.undp.org/data-center/documentation-and-downloads (last verified 14 Aug 2022) |
| Journal impact factor | Journal impact factor from Clarivate’s Journal Citation Reports for the mean study year, averaged across journals included in the study. See Online Methods for more details. See Supplementary Data S1 for a full list of journals used within the studies. |
| Journal country | Used in journal nationalism dataset. Where the journal is/was located according to the manuscript (if mentioned) or Clarivate’s Journal Citation Reports (if not mentioned) |
| Journals submitted to | Used in the number of submissions database. The number of journals a manuscript was submitted to prior to publication. |
| Language | Used as a column in data at the individual country level. Primary language for the country of the author’s institutional affiliation. Assigned using the same scheme reported elsewhere. |
| last\_author\_country | Used in the number of submissions database. Country of the last author’s affiliation on the manuscript. |
| last\_author\_gender | Used in the number of submissions database. Assumed gender of the last author. |
| Last\_continent | Used in the number of submissions database. Continent matching the country of the last author’s affiliation on the manuscript. |
| Last\_HDI2012 | Used in the number of submissions database. Country Human Development Index matching the country of the last author’s affiliation on the manuscript during the mean study year (2012). |
| Last\_language | Used in the number of submissions database. Country primary language matching the country of the last author’s affiliation on the manuscript. |
| Last\_language\_similar\_HDI | Used in the number of submissions database. Country primary language matching the country of the last author’s affiliation on the manuscript. This column only includes rows whose author’s country was within 0.10 HDI points of the highest HDI country in that dataset to control for HDI possibly confounding language effects. |
| Major revision | Used in reviewer scores data. The number of manuscripts recommended for “major revision” by reviewers. Studies were conducted on journals with a wide variety of actual possible categories/labels, so details on the actual labels from the original studies are in “Notes” |
| Max value | Used in reviewer scores datasets. The maximum possible value for a review score in that study. |
| Max.JIF.submitted.to | Used in the number of submissions database. Highest journal impact factor of all journals the manuscript was submitted to prior to publication. |
| Min value | Used in reviewer scores datasets. The minimum possible value for a review score in that study. |
| Minor revision | Used in reviewer scores datasets. The number of manuscripts recommended for “minor revision” by reviewers. Studies were conducted on journals with a wide variety of actual possible categories/labels, so details on the actual labels from the original studies are in “Notes” |
| N.reviewers | Average number of reviewers per manuscript for that author position/demographic category for that study. Only in the number of reviewers files. |
| N.revisions | Average number of revisions per manuscript for that author position/demographic category for that study. Only in the number of revisions files. |
| Notes | Any pertinent notes on how data were calculated or caveats in their interpretation/use |
| Reject | Used in reviewer scores datasets. The number of manuscripts recommended for “rejection” by reviewers. Studies were conducted on journals with a wide variety of actual possible categories/labels, so details on the actual labels from the original studies are in “Notes” |
| Reject with invitation to resubmit | Used in reviewer scores datasets. The number of manuscripts recommended for “reject with invitation to resubmit” by reviewers. Studies were conducted on journals with a wide variety of actual possible categories/labels, so details on the actual labels from the original studies are in “Notes” |
| Rejected | Number of manuscripts rejected at that review stage |
| Reviewer country | Used in reviewer homophily dataset where data were available at the individual country level. Country where the reviewer’s affiliation was in. |
| Reviewer demographic | Reviewer demographic category following same labelling scheme as author demographic category. There are two novel labelling schemes for reviewer demographics pertaining to assumed gender and country homophily.  Country: Match = reviewer’s and author’s affiliation were based in the same country; Not match = reviewer’s and author’s affiliation were not based in the same country  Gender\_1: studies used the individual reviewer as the unit of replication and summarized scores for each reviewer/author interaction. Labelling uses “Male” or “Z\_Female” in line with Demographic.category.  Gender\_2: study used the manuscript as the unit of replication. “1+ Female” = 1 or more reviewers were assigned female; “All Male” = all reviewers were assigned male |
| Reviewer ordinal scores | Reviewer scores, converted to a weighted continuous scale if provided by the study as ordinal (e.g., 1 = accept, 2 = minor revision, 3 = major revision, 4 = reject). Occasionally studies provided reviewer scores on already continuous scales like priority rankings from 0–100. See Supplementary Table 2, reproduced above, for details. |
| Reviewer\_rescaled | “Reviewer ordinal scores” re-scaled to range from 0 to 1 by taking ([value] – [min value])/([max value] – [min value]). |
| Stage | Peer review stage: see Supplementary Table 2 (reproduced above) and Supplementary Fig. 1 for definitions. |
| Study | Study name carried through all files. Matched to study citation information in “Dataset S1 Studies screened for inclusion in meta-analysis.csv” |
| Study type | Experiment – in this context, random assignment to single-blind/double-blind review used by Alam et al. (2011). Observational – study uses data from real peer review without manipulation.  Observational before/after – used for Jagsi et al. (2014) where they examined peer review outcomes by author demographics before and after implementation of double-blind review in a formerly single journal.  Questionnaire survey – author questionnaire survey (only available for Fox and Paine (2019) number submissions data) |
| Subcategory | Biological Sciences sub-category for journals included within that study:  Broad = covered multiple fields but publishes in the biological sciences like *Nature*  Health sciences = medical journals  Life sciences in this category was further broken down into Ecology & Evolution (defined using the list of journals used in our journal policy analysis except that Tier 2 journals were labelled as Broad) and Other Biological Sciences (non-Ecology & Evolution journals that were defined as Life Sciences).  We used the majority type of journal if more than one was used in that study. In the case of Squazzoni et al. (2021), we maintained the health sciences/life sciences breaks provided in their original dataset since the study covered >100,000 manuscripts and 79 journals and could be broken up easily. |
| Time data | Average time in days to complete that peer review stage for the author position/demographic category that row corresponds to |
| Total manuscripts | Number of manuscripts corresponding to that row of data. In reviewer scores datasets, this is the sample size underlying the ordinal scores. It may include multiple author\*reviewer interactions on the same manuscript. |
| Went through to next stage | If manuscript was not rejected at that review stage. This is equivalent to accepted for final decision and overall decision. |
| Year | Mean year data were collected from journals for that study. Rescaled in the R code so that the earliest year across studies = 0. This was done to improve convergence. |

**REFERENCES**

Alam, M., Kim, N.A., Havey, J., Rademaker, A., Ratner, D., Tregre, B., *et al.* (2011). Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: A randomized multi-rater study. *Br. J. Dermatol.*, 165, 563–567.

Burns, C.S. & Fox, C.W. (2017). Language and socioeconomics predict geographic variation in peer review outcomes at an ecology journal. *Scientometrics*, 113, 1113–1127.

Fox, C.W., Burns, C.S., Muncy, A.D. & Meyer, J.A. (2016). Gender differences in patterns of authorship do not affect peer review outcomes at an ecology journal. *Funct. Ecol.*, 30, 126–139.

Fox, C.W. & Paine, C.E.T. (2019). Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. *Ecol. Evol.*, 9, 3599–3619.

Goldstone, K., Edgley, C., Mehta, S. & Leslie, K. (2020). Peer review for the Canadian Journal of Anesthesia in 2016 and 2017: a retrospective analysis by reviewer and author gender. *Can. J. Anesth.*, 67, 336–342.

Jagsi, R., Bennett, K.E., Griffith, K.A., Decastro, R., Grace, C., Holliday, E., *et al.* (2014). Attitudes toward blinding of peer review and perceptions of efficacy within a small biomedical specialty. *Int. J. Radiat. Oncol. Biol. Phys.*, 89, 940–946.

Opthof, T., Coronel, R. & Janse, M.J. (2002). The significance of the peer review process against the background of bias: Priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovasc. Res.*, 56, 339–346.

Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., Willis, M., *et al.* (2021). Peer review and gender bias: A study on 145 scholarly journals. *Sci. Adv.*, 7.

Zhang, X. (2012). Effect of reviewer’s origin on peer review: China vs. non-China. *Learn. Publ.*, 25, 265–270.