Journal of the American Heart Association

SYSTEMATIC REVIEW AND META-ANALYSIS

Effects of Experimental Interventions to Improve the Biomedical Peer-Review Process: A Systematic Review and Meta-Analysis

Mario Gaudino, MD; N. Bryce Robinson, MD; Antonino Di Franco, MD; Irbaz Hameed, MD; Ajita Naik, MD; Michelle Demetres, MLIS; Leonard N. Girardi, MD; Giacomo Frati, MD; Stephen E. Fremes, MD, MSc; Giuseppe Biondi-Zoccai, MD

BACKGROUND: Quality of the peer-review process has been tested only in small studies. We describe and summarize the randomized trials that investigated interventions aimed at improving peer-review process of biomedical manuscripts.

METHODS AND RESULTS: All randomized trials comparing different peer-review interventions at author-, reviewer-, and/or editor-level were included. Differences between traditional and intervention-modified peer-review processes were pooled as standardized mean difference (SMD) in quality based on the definitions used in the individual studies. Main outcomes assessed were quality and duration of the peer-review process. Five-hundred and seventy-five studies were retrieved, eventually yielding 24 randomized trials. Eight studies evaluated the effect of interventions at author-level, 16 at reviewer-level, and 3 at editor-level. Three studies investigated interventions at multiple levels. The effects of the interventions were reported as mean change in review quality, duration of the peer-review process, acceptance/rejection rate, manuscript quality, and number of errors detected in 13, 11, 5, 4, and 3 studies, respectively. At network meta-analysis, reviewer-level interventions were associated with a significant improvement in review quality (SMD, 0.20 [0.06 to 0.33]), at the cost of increased duration of the review process (SMD, 0.15 [0.01 to 0.29]), except for reviewer blinding. Author- and editor-level interventions did not significantly impact peer-review quality and duration (respectively, SMD, 0.17 [−0.16 to 0.51] and SMD, 0.19 [−0.40 to 0.79] for quality, and SMD, 0.17 [−0.16 to 0.51] and SMD, 0.19 [−0.40 to 0.79] for duration).

CONCLUSIONS: Modifications of the traditional peer-review process at reviewer-level are associated with improved quality, at the price of longer duration. Further studies are needed.

REGISTRATION: URL: https://www.crd.york.ac.uk/prospero; Unique identifier: CRD42020187910.

Key Words: network meta-analysis ■ peer-review ■ review quality

See Editorial by London

Peer-review is the gold standard for reviewing scientific contributions.[1] This process has often been criticized as being poorly evidence-based,[2] time-consuming, expensive, and open to biases.[3] These limitations appear even more evident in the current COVID-19 pandemic, with a dire need for timely information which is at odds with the requirements of time consuming peer-review.

Over the years, several efforts have been made to improve the quality of peer-review.[4] Few of these,

Correspondence to: Mario Gaudino, MD, Department of Cardiothoracic Surgery, Weill Cornell Medicine, 525 E. 68th St, New York, NY 10065. E-mail: mfg9004@med.cornell.edu

Supplementary Material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.120.019903

For Sources of Funding and Disclosures, see page 13.

CLINICAL PERSPECTIVE

What Is New?

- While most interventions on the traditional peer-review process do not significantly improve its quality, modifications at reviewers-level are associated with improved quality at the price of a longer duration of the review process.

What Are the Clinical Implications?

- Further investigation into interventions aimed at improving the peer-review process is needed.

Nonstandard Abbreviations and Acronyms

MD	mean difference
SMD	standardized mean difference

however, have led to unequivocal and significant improvements.[5]

In this systematic review and network meta-analysis we aimed to quantitatively evaluate the effect of the different interventions tested in randomized trials focusing on improving the quality or efficiency of the peer-review process.

METHODS

A systematic review and meta-analysis of all published or registered randomized trials assessing interventions aimed at improving quality of the biomedical peer-review process was performed, after formal design disclosure (PROSPERO ID: CRD42020187910). The data that support the findings of this study are available from the corresponding author upon reasonable request.

Search Strategy

A medical librarian (M.D.) performed a comprehensive search to identify contemporary randomized trials on peer-review (no language restrictions). Searches were run on December 2019 in Ovid MEDLINE and updated on June 12, 2020. The full search strategy is available in Table S1.

Study Selection and Data Extraction

Two independent reviewers (N.B.R. and I.H.) screened retrieved studies; discrepancies were resolved by the senior author (G.B.Z.). Titles and abstracts were reviewed against predefined inclusion/exclusion criteria. Articles were considered for inclusion if they were randomized trials reporting comparisons between different peer-review interventions at author-, reviewer-, and/or editor-level, aimed at improving quality of the peer-review process by exploring at least one of the following outcomes: acceptance/rejection rate, quality of the manuscript, quality of the review, duration of the peer-review process, number of errors detected. Case reports, conference presentations, editorials, expert opinions, and studies not comparing review processes were excluded.

Full texts of the selected studies were examined for a second round of eligibility screening. Reference lists for articles selected for inclusion were also searched for relevant articles (backward snowballing). All studies were reviewed by 2 independent investigators and discrepancies were resolved by the senior author. The full Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram outlining the study selection process is available in Figure S1.

Interventions were classified based on the process level at which they operated (author-level, reviewer-level, editor-level). The following variables were extracted for each study: study level data (sample size, year, country of origin, journal), interventions tested, main outcomes assessed, level of intervention (author-, reviewer-, editor- level), assessors of review quality, timing of the assessment, assessment method, main findings, and summary of the effects of the interventions. For studies with multiple interventions, data were separately collected for each intervention. Two investigators performed data extraction independently; the extracted data were verified by a third investigator for accuracy.

The quality of the included studies was assessed using the Cochrane Collaboration's Tool for assessing Risk of Bias in randomized trials (Table S2).

Network Meta-Analysis

The main outcome assessed was the quality of the peer-review process. The differences between traditional and intervention-modified peer-review were pooled as standardized mean difference (SMD) in quality based on the definitions used in the individual studies. Duration of the peer-reviewing process, defined as time-to-decision, was also compared. Random-effects network meta-analysis was performed using the generic inverse variance method with the netmeta statistical package in R with the study control groups serving as the reference. The Cochran's Q statistic was used to assess inconsistency. Rank scores with probability ranks of different treatment groups were calculated.

Small study effects and publication bias were assessed with comparison-adjusted funnel plots and regression tests. Leave-one-out sensitivity analysis and a sensitivity analysis based on fixed effect methods were also performed. Statistical significance was set at the 2-tailed 0.05 level, without multiplicity adjustment.

All statistical analyses were performed using R (version 3.5.2, R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Description of the Studies

Searches across the chosen databases retrieved 622 studies. After results were de-duplicated, a total of 575 studies were retrieved of which 24 studies met inclusion criteria (Table 1).[6–29] There were 13 studies originating from the United States, 6 from the United Kingdom, 3 from Spain, 1 from India, and 1 from Denmark. There were 9 studies published before the year 2000 and 15 after 2000.

Level of Interventions and Outcomes Assessed

There were 8 studies evaluating the effect of interventions at the author-level, 16 at the reviewer-level, and 3 at the editor-level. Three studies evaluated interventions at more than one level (Tables 1 and 2).

The outcomes assessed were reported as mean change in review quality in 13 studies (evaluated by means of either a 5-point scale [11 studies], a predesigned form scoring 1–100 [1 study], or an 8-item review quality instrument [1 study]), duration of the peer-review process in 11 studies, acceptance/rejection rate in 5 studies, manuscript quality in 4 studies (by means of either a 5-point scale [3 studies], or the Modified Manuscript Quality Assessment Instrument [1 study]), and number of errors detected in 3 studies (further details are provided in Table 1).

Author-Based Interventions

Eight studies investigated the impact of author-level interventions on the quality of the review; van Rooyen analyzed 467 manuscripts submitted to *British Medical Journal* where one reviewer was blinded to author identity, and the other was not. The authors found no significant differences in review quality between the groups, as measured by mean total quality score (mean difference [MD], 0.02; 95% CI, −0.11 to 0.14); however, the author highlighted that their results were likely not generalizable to other journals.[25]

Fisher et al assigned 57 manuscripts to reviewers blinded or unblinded to author identity and found that while there was no difference in mean rating scores (scores 1 to 5 [1=accept; 5=reject]), but unblinded reviewers gave higher priority scores to authors with more published articles (Spearman rank correlation coefficient [r]=−0.45 for blinded group versus r=−0.14 for unblinded group).[13]

In a similar study, Alam et al found no differences in the rates of acceptance (37.5% versus 32.5%), revision (48.75% versus 47.4%), or rejection (13.75% versus 20.0%) between blinded and unblinded reviewers (P=0.32).[6]

Godlee and colleagues performed a randomized trial in which reviewers were given a paper accepted to publication with 8 known errors. The reviewers were either blinded or unblinded to authors identity, and then asked to either sign or not sign their comments. Neither blinding reviewers to the authors and origin of the paper nor requiring them to sign their reports had effect on rate of detection of errors. However, reviewers who were blinded to author identity were more likely to recommend acceptance (odds ratio (OR), 0.5; 95% CI, 0.3–1.0).[14]

Okike et al evaluated the impact of blinding of reviewers on acceptance rate and identification of 5 known errors. They found that both recommendation for acceptance (MD, 1.28; 95% CI, 1.06–1.39; P=0.02) and attribution of higher scores to the manuscript (MD, 1.35; 95% CI, 0.56–2.13; P<0.001) were more likely in the unblinded group.[21]

John et al found no effect on the quality of the review by providing the authors' conflict of interest disclosures to the reviewers (MD, 0.04; 95% CI, −0.05 to 0.14).[16]

McNutt et al reported that blinding improved review quality (MD, 0.41; P=0.007); no difference was reported in terms of acceptance rate and time to review.[19]

Justice et al found no difference in review quality in a trial of 118 manuscripts randomized to a control group (where journals followed their usual practice) or to an intervention group (where one reviewer knew authors identities, and the other was blinded).[18]

Reviewer-Based Interventions

Sixteen studies tested the impact of reviewer-level interventions. van Rooyen and colleagues in 2 separate analyses reported that revealing reviewer identity did not significantly impact the quality of review, although it increased the amount of time for the review to be written.[26,27]

Godlee et al found that revealing reviewers' identity did not affect the rejection rate (OR, 0.5; 95% CI, 0.3–1.0)[14] and McNutt and associates reported that revealing the reviewer's identity did not change the quality of the reviews.[19]

Walsh et al found a significant difference in mean quality between blinded and unblinded reviewers (3.35 versus 3.14, P=0.02), but the small absolute difference

Table 1. Randomized Trials Included in the Analysis

Author, Y	Country of Origin	Journal	Sample Size	Interventions	Main Outcomes Assessed	Level of the Intervention	Assessors	Timing of the Assessment	Assessment Modalities	Main Results	Summary of the Effects of the Interventions
Alam et al, 2011[6]	United States	*British Journal of Dermatology*	40 manuscripts 4 reviewers	Blinding to author identity	Primary outcome: Acceptance rate Secondary outcome: word count of the narrative portion of the reviewer form	Author	Reviewers	At peer-review	Electronic forms with checkboxes on manuscript quality, a series of checkboxes for suggesting disposition of the manuscript (ie, accept, minor revisions, major revisions, reject), and narrative sections for comments to the editors and authors, respectively	No difference in acceptance rate (37.5% vs 32.5%), revise rate (48.75% vs 47.4%), and rejection rate (13.75% vs 20.0%) between blinded and unblinded reviewers (P=0.32) No difference in word count between unblinded and blinded reviews	No effect
Arnau et al, 2003[7]	Spain	*Medicina Clinica*	43 manuscripts	Addition of a statistical review to a clinical review	Quality of the manuscript	Reviewer	Statistician	Before and after peer-review	Modified Manuscript Quality Assessment Instrument	MD, 1.35; 95% CI, −0.45 to 3.16, P=0.13)	No effect
Callaham and Schriger, 2002[8]	United States	*Annals of Emergency Medicine*	Study 1: 50 reviewers Study 2: 22 reviewers	Training	Quality of the review	Reviewer	Editors	For 2 y after workshop attendance	1- to 5-point scale	Study 1: MD, 0.11 (95% CI, −0.25 to 0.48) in the control vs 0.10 (95% CI, −0.20 to 0.39) in the intervention group Study 2: MD −0.10 (95% CI, −0.49 to 0.29) in the control vs 0.06 (95% CI, −0.34 to 0.23) in the intervention group	No effect
Callaham et al, 2002[9]	United States	*JAMA*	Study 1: 35 reviewers Study 2: 95 reviewers	Written feedback to reviewers from the editors	Quality of the review	Editor	Editor	After peer-review	1- to 5-point scale	Study 1: MD, 0.16 (95% CI, 0.26 to 0.58) in the control vs −0.13 (95% CI, −0.29 to 0.23) in the intervention group Study 2: MD, 0.12 (95% CI, −0.20 to 0.26) in the control vs 0.06 (95% CI, −0.19 to 0.31) in the intervention group	No effect

(*Continued*)

Table 1. Continued

Author, Y	Country of Origin	Journal	Sample Size	Interventions	Main Outcomes Assessed	Level of the Intervention	Assessors	Timing of the Assessment	Assessment Modalities	Main Results	Summary of the Effects of the Interventions
Cobo et al, 2007[10]	Spain	Plos One	30 Clinical review+Statistical reviewer 28 Clinical review+Guidelines/ Checklists	Addition of a statistical reviewer Suggestion to use guidelines/ checklists to clinical reviewers	Quality of the manuscript	Reviewer	Authors of the paper	After peer-review	5-point Likert scale	Addition of a statistical reviewer: mean rating change 5.5 (95% CI, 4.3 to 6.7) Suggestion to use guidelines/checklists: mean rating change 0.9 (95% CI, −0.3 to 2.1)	Addition of a statistical reviewer: positive effect Suggestion to use guidelines/ checklists: no effect
Cobo et al, 2011[11]	Spain	*BMJ*	92 manuscripts	Additional review by the editor based on reporting guidelines	Primary outcome: Quality of the manuscript Secondary outcome: average of all pertinent items (after excluding specific items that did not apply to the study)	Editor	Junior statisticians	After second editorial decision following peer-review	5-point Likert scale	Primary outcome: Comparison as allocated, MD, 0.25; 95% CI, −0.05 to 0.54; as reviewed, MD 0.33; 95% CI, 0.03 to 0.63 Secondary outcome: Comparison as allocated, MD 0.11; 95% CI, (−0.01 to 0.22); as reviewed, MD, 0.15; 95% CI, 0.04 to 0.27)	Positive effect
Das Sinha et al, 1999[12]	India	*National Medical Journal of India*	78 manuscripts	Review exchange among Indian and non-Indian reviewers	Quality of the review Time taken to complete review	Reviewer	Editors	After peer-review	Pre-designed form scoring up to 100 Turnaround time	Being informed that reviews would be exchanged did not affect the quality of reviews by non-Indians (54.8 exchanged vs 58.4 non-exchanged) or of reviews by Indians (50.0 exchanged vs 47.3 non-exchanged) Non-Indian reviewers scored higher than Indians (56.7 vs 48.6, P<0.001) Non-Indians took the same amount of time as Indians to return their reviews	No effect
Fisher et al, 1994[13]	United States	*JAMA*	57 manuscripts 112 blinded reviewers 108 non-blinded reviewers	Blinding to author identity	Rejection rate	Author	Reviewers and Editors	At peer-review	Rating scores of 1 to 5 (1, accept; 5, reject)	Outright rejection: 30% vs 21% for blinded vs un-blinded reviewers Acceptance requiring revisions: 28% vs 36% for blinded vs un-blinded reviewers	Blinded reviewers provided more unbiased reviews

(*Continued*)

Table 1. Continued

Author, Y	Country of Origin	Journal	Sample Size	Interventions	Main Outcomes Assessed	Level of the Intervention	Assessors	Timing of the Assessment	Assessment Modalities	Main Results	Summary of the Effects of the Interventions
Godlee et al, 1998[14]	UK	*JAMA*	420 reviewers allocated to 4 groups	Blinding to author identity Revealing reviewer's identity	No. of errors detected Rejection rate	Author and Reviewer	Authors of the paper	After peer-review	No. of weakness detected out of 8 total weaknesses purposely introduced in design, analysis, or interpretation of the paper Proportion recommending rejection	Mean number of errors detected did not differ among groups Blinded reviewers were less likely to reject (odds ratio, 0.5; 95% CI, 0.3 to 1.0)	No effect in detection of errors Blinded reviewers less likely to reject
Houry et al, 2012[15]	United States	*BMC Medical Education*	24 mentored reviewers 22 controls	Training	Quality of the review	Reviewer	Mentors	After peer-review	1- to 5-point quality scale	Size effect 0.1 (95% CI, −0.4 to 0.6)	No effect
John et al, 2019[16]	United States	*BMJ*	1480 manuscripts 838 reviewers	Unblinding to authors' disclosures	Quality of the manuscript	Author	Reviewers	After peer-review	1-to 5-point quality scale	Mean desirability score 2.70 (SD 1.11) out of 5 in the control group vs 2.74 (1.13) out of 5 in the intervention group; MD 0.04 (95% CI, −0.05 to 0.14)	No effect
Johnston et al, 2007[17]	United States	*Annals of Neurology*	88 manuscripts undergoing traditional review 263 manuscripts undergoing early screening	Early screening	Time to final decision and number of required reviews for each manuscript	Editor	Editors	Before peer-review	Editors' judgement	Time to final decision: 48 d vs 18 d for traditional vs early screening (P<0.0001) Mean number of require reviews for each manuscript: 2.3 vs 0.7 for traditional vs early screening (P<0.0001)	Positive effect
Justice et al, 1998[18]	United States	*JAMA*	118 manuscripts	Blinding to author identity	Quality of the review	Author	Editors and Authors	After peer-review	5-point Likert scale	MD, 0.1 (95% CI, −0.2 to 0.4) in blinded vs −0.1 (95% CI, −0.5 to 0.4) in unblinded reviews	No effect
McNutt et al, 1990[19]	United States	*JAMA*	127 manuscripts	Blinding to author identity Unblinding reviewer's identity	Quality of the review Time taken to complete review Reviewer's recommendation on publication	Author and Reviewer	Editors and Authors	After peer-review	5-point scale	Quality of the review: MD, 0.41; P=0.007 for blinded vs unblinded reviews No association between unblinding reviewers' identity and quality of the review No significant differences between groups with regard to recommendation to publication or time taken to review	Positive effect on review quality No effect on acceptance rate and time to review

(*Continued*)

Table 1. Continued

Author, Y	Country of Origin	Journal	Sample Size	Interventions	Main Outcomes Assessed	Level of the Intervention	Assessors	Timing of the Assessment	Assessment Modalities	Main Results	Summary of the Effects of the Interventions
Neuhauser and Koran, 1989[20]	United States	*Medical Care*	95 manuscripts	Calling reviewers before review	Time to final decision	Reviewer	Authors of the paper	After peer-review	Turnaround time	Mean turnaround time: 44.2 d vs 37.7 d for intervention vs control group	Negative effect
Okike et al, 2016[21]	United States	*JAMA*	119 reviewers	Blinding to author identity	Primary outcome: Rejection rate Secondary outcomes: number of intentionally placed errors detected and quality scores for the Methods	Author	Reviewers	At peer-review	Proportion recommending rejection Reviewer score (range, 0–10) No. of errors detected (max 5)	Unblinded reviewers more likely to recommend acceptance (relative risk, 1.28; 95% CI, 1.06 to 1.39, P=0.02) and give higher scores (MD, 1.35; 95% CI, 0.56 to 2.13; P<0.001)	Unblinded reviewers more likely to recommend acceptance and give higher scores
Pitkin and Burmeister, 2002[22]	United States	*JAMA*	283 manuscripts	Asking reviewer to accept before assignment of manuscript	Time taken to complete the review Quality of the review	Reviewer	Editors	After peer-review	5-point scale	Mean time to file a review once manuscript was mailed: 21.0 vs 25.0 d for intervention vs control, respectively (P<0.001) Overall time to receipt of review: 24.7 vs 25.9 d, for intervention vs control, respectively P=0.19) No difference observed in review quality (P=0.39)	Positive effect on turnaround time No effect on total manuscript processing time nor on review quality
Provenzale, 2020[23]	United States	*American Journal of Roentgenology*	201 manuscripts	Assigning reviewers 1 d vs 3 d to accept invitation	Time taken to complete the review	Reviewer	Authors of the paper	After peer-review	Turnaround time	Mean turnaround time: 27.9 d vs 31.5 d in the 1-d vs 3-d invitation groups, respectively (P=0.04)	Positive effect

(*Continued*)

Table 1. Continued

Author, Y	Country of Origin	Journal	Sample Size	Interventions	Main Outcomes Assessed	Level of the Intervention	Assessors	Timing of the Assessment	Assessment Modalities	Main Results	Summary of the Effects of the Interventions
Schroter et al, 2004[24]	UK	*BMJ*	609 reviewers	Training (workshop vs self-taught)	Quality of the review No. of intentionally placed errors detected Time taken to complete the review Rejection rate	Reviewer	Authors of the paper	After peer-review	8-item review quality instrument No. of errors detected Turnaround time Proportion recommending rejection	Quality of the review: self-taught group scored higher than controls (MD, 0.29; 95% CI, −0.14 to 0.44; P=0.001) as well as face-to-face group (MD, 0.16; 95% CI, 0.02 to 0.3; P=0.025) but the difference was not of editorial significance and was not maintained in the long term No. of errors detected: both intervention groups identified significantly more major errors after training than did the control group (3.14 and 2.96 vs 2.13; P<0.001) Training had no impact on the time taken to review the papers but was associated with an increased likelihood of recommending rejection (92% and 84% vs 76%; P=0.002)	Positive effects only in the short term
van Rooyen et al, 1998[25]	UK	*BMJ*	467 manuscripts	Blinding to author identity and/or revealing the reviewer's identity to a coreviewer	Quality of the review Time taken to complete the review	Author and Reviewer	Editors	After peer-review	5-point Likert scale Turnaround time	Quality of the review: MD, 0.02 (95% CI, −0.11 to 0.14) for blinded vs non-blinded MD, 0.16 (95% CI, −0.29 to −0.02) for revealing vs not revealing reviewer's identity Time taken to complete the review: no significant difference between blinding vs unblinding; no significant difference between revealing vs not revealing reviewer's identity	No effect

(*Continued*)

Table 1. Continued

Author, Y	Country of Origin	Journal	Sample Size	Interventions	Main Outcomes Assessed	Level of the Intervention	Assessors	Timing of the Assessment	Assessment Modalities	Main Results	Summary of the Effects of the Interventions
van Rooyen et al, 1999[26]	UK	*BMJ*	113 manuscripts	Revealing reviewer's identity	Primary outcome: Quality of the review Secondary outcomes: time taken to complete the review and recommendation of publication	Reviewer	Editors and corresponding authors	After peer-review	5-point Likert scale Turnaround time Proportion recommending rejection	Quality score 3.06±0.72 vs 3.09±0.68 for controls vs intervention (*P*=0.68), MD, 0.03; 95% CI, −0.19 to 0.12) No significant difference in the recommendation of publication or time taken to review the paper.	No effect
van Rooyen et al, 2010[27]	UK	*BMJ*	471 reviewers 12 editors	Revealing reviewer's identity	Quality of the review Time taken to complete the review Reviewer's recommendation of publication	Reviewer	Editors and corresponding authors	After peer-review	5-point Likert scale Turnaround time Proportion recommending rejection	No significant difference on quality of the review between the intervention and control groups (MD for editors, 0.04; 95% CI, −0.09 to 0.17; MD for authors, 0.06; 95% CI, −0.09 to 0.20) Reviewers in the intervention group took significantly longer to review (MD, 25 min; 95% CI, 3.0 to 47.0) No significant difference on the likelihood that reviewers would recommend acceptance between the intervention and control groups (MD, 4%; 95% CI, −5% to 13%) or that the paper would ultimately be accepted for publication (MD, −5%; 95% CI, −13% to 3%)	No effect on quality of the review Negative effect on the amount of time to write a review
Vinther et al, 2012[28]	Denmark	*Danish Medical Journal*	364 reviews	Revealing reviewer's identity	Quality of the review	Reviewer	Editors	After peer-review	5-point scale	Quality score 3.34 vs 3.28 for intervention vs controls (*P*=0.51)	No effect
Walsh et al, 2000[29]	UK	*British Journal of Psychiatry*	245 reviewers	Revealing reviewer's identity	Quality of the review Time taken to complete the review	Reviewer	Editors	After peer-review	5-point scale	Quality score 3.35 vs 3.14 for intervention vs controls (*P*=0.02)	Positive effect

BMJ indicates *British Medical Journal*; *JAMA*, *Journal of the American Medical Association*; and MD, mean difference.

Table 2. Summary of Level, Types, and Outcome of the Tested Interventions on the Peer-Review Process

Level of the Intervention	Intervention	Effect on the Peer-Review Process in Single Studies	Summary of the Effect on the Peer-Review Process
Editor	Early screening	+ (Johnston et al, 2007)[17]	+
Author	Blinding authors' identity	± (Alam et al, 2011)[6] + (Fisher et al, 1994)[13] + (Godlee et al, 1998)[14] ± (Justice et al, 1998)[18] + (McNutt et al, 1990)[19] + (Okike et al, 2016)[21] ± (van Rooyen et al, 1998)[25]	+
	Unblinding authors' disclosures	± (John et al, 2019)[16]	±
Reviewer	Geographic exchange	± (Das Sinha et al, 1999)[12]	±
	Prior reviewer contact	– (Neuhauser and Koran, 1989)[20] +/± (Pitkin and Burmeister, 2002)[22] + (Provenzale, 2020)[23]	+
	Guidelines/training	± (Callaham and Schriger, 2002)[8] ± (Cobo et al, 2007)[10] ± (Houry et al, 2012)[15] + (Schroter et al, 2004)[24]	+
	Revealing reviewer's identity	+ (Godlee et al, 1998)[14] ± (McNutt et al, 1990)[19] ± (van Rooyen et al, 1998)[25] ± (van Rooyen et al, 1999)[26] ±/– (van Rooyen et al, 2010)[27] ± (Vinther et al, 2012)[28] + (Walsh et al, 2000)[29]	±
	Statistical review	± (Arnau et al, 2003)[7] + (Cobo et al, 2007)[10]	+
Editor	Editorial review	+ (Cobo et al, 2011)[11]	+
	Feedback from the editor	± (Callaham et al, 2002)[9]	±

JAMA indicates *Journal of the American Medical Association*.

does not seem to support a clear advantage of one approach over another.[29] Das Sinha reported no difference in mean review score when reviewers were informed that a copy of their comments would be sent to other reviewers working on the same manuscript[12] and Vinther and colleagues found no difference between blinded and unblinded reviewers (mean quality score 3.34 for unblinded reviewers versus 3.29 for blinded reviewers, P=0.51).[28]

Schroter et al investigated the impact of reviewer training. Reviewers underwent either a face-to-face training, a self-taught module, or no training and were sent 3 papers with deliberate errors added. A slight improvement in quality was seen in the self-taught group (MD, 0.29; 95% CI, 0.14–0.44; P=0.001) and face-to-face group (MD, 0.16; 95% CI, 0.02–0.3; P=0.025) when compared with controls. This improvement, however, was transient, as disappeared upon review of the third paper.[24] Callaham and Schriger invited reviewers to attend a 4-hour formal workshop. While most (81%) found the workshop helpful and 85% of attendees felt that the quality of their review would improve, the authors did not find a significant difference in the mean quality of review between attendees and controls (MD, 0.11; 95% CI, −0.25 to 0.48 for controls versus MD, 0.10; 95% CI, −0.20 to 0.39 for intervention group).[8]

Houry et al tested the efficacy of pairing new reviewers with senior reviewers and found that the quality of the review did improve significantly (effect size, 0.1; 95% CI, −0.4 to 0.6).[15]

Three studies evaluated interventions at reviewer-level aimed at decreasing turnaround time. Two of these studies investigated the impact of contacting reviewers before manuscript assignment. Neuhauser and Koran found that this strategy increased turnaround time (from 37.7 to 44.2 days)[20] while, Pitkin and Burmeister found a significant reduction in review turnaround time (21.0 days versus 25.0 days, P<0.001) but not in the overall manuscript processing time (24.7 days versus 25.9 days, P=0.19), in large part because of the high rate (15%) of reviewers who declined in the ask-first group.[22] Provenzale and co-authors found a significant decrease in turnaround time when reviewers were given 1 instead of 3 days to accept the invitation to review (total turnaround time 27.9 days in 1-day decision group versus 31.5 days in 3-day decision group, P=0.04).[23]

Two studies investigated the impact of the addition of a statistical reviewer on review quality. Cobo et al found that addition of a statistical reviewer improved the quality of the review (MD, 5.5; 95% CI, 4.3–6.7),[10] while, Arnau and colleagues reported no effect (MD, 1.35; 95% CI, −0.45 to 3.16; *P*=0.13), although in a per-protocol analysis, a significant difference in favor of the group with statistical reviewers was found (MD, 1.96; 95% CI, 0.25–3.67; *P*=0.026).[7]

Editor-Based Interventions

Three studies investigated the impact of editor-level interventions. Callaham et al asked editors to give written feedback to poor-quality and average-quality reviewers and found that the review quality did not significantly change, (MD, −0.13; 95% CI, −0.49 to 0.23 in the poor quality and 0.06, 95% CI, −0.19 to 0.31 in the average quality group).[9]

Cobo and associates investigated the use of checklists such as CONSORT and STROBE and found that it improves manuscript quality, although the observed effect was small (MD, 0.33; 95% CI, 0.03–0.63 for the comparison "as reviewed").[11] Johnston et al tested the effect of in-house editorial screening before external review and found that it significantly decreased review time (from 48 days to 18 days, *P*<0.001).[17]

Network Meta-Analysis

Twenty-four studies were included in the network meta-analysis for the outcome of peer-review quality (Figure 1, Figure S2, and Tables S3 and S4). Compared with traditional process, reviewer-level interventions were associated with a significant improvement in the quality of peer review (SMD, 0.20; 95% CI, 0.06–0.33). There was no significant improvement associated with author- (SMD, 0.10; 95% CI, −0.11 to 0.30) and editor-level interventions (SMD, 0.01; 95% CI, −0.32 to 0.34) (Table 3). Reviewer-level interventions ranked as the best intervention (rank score for reviewer-level 0.88 versus 0.57 for author-level, and 0.34 for editor level) (Figure 1).

The level of evidence was high for all comparisons (Table S2). Heterogeneity/inconsistency and netsplit analyses are shown in Tables S3 and S4. Egger test for a regression intercept indicated no evidence of publication bias (*P*=0.18) (Figure S3). Leave-one-out analysis confirmed the solidity of the results (Figure S4). Sensitivity analysis based on fixed-effect methods confirmed the main analysis (Figures S5 and S6, Tables S5 and S6).

The impact of the interventions at different levels (author-, reviewer-, and editor-level) on the duration of the peer-review process was also tested (Figure 2, Figures S7 through S9, Tables S7 through S9). Interventions at reviewer-level were associated with a significant increase in the length of the peer-review process (SMD, 0.15; 95% CI, 0.01–0.29), while author- and editor-level interventions were not (SMD, 0.17; 95% CI, −0.16 to 0.51 and SMD, 0.19; 95% CI, −0.40 to 0.79, respectively) (Table 3). Sensitivity analysis confirmed these results (Figures S10 and S11, Tables S10 and S11).

Among the different reviewer-level interventions tested, unblinding reviewer's identity was the only modality that did not significantly impact duration of the peer-review process (SMD, 0.01; 95% CI, −0.17 to 0.19) (Figures S12 through S15, Tables S12 through S14).

DISCUSSION

In the present quantitative synthesis, we found that among the different interventions proposed to improve the process of peer-review, those directed at reviewer level were associated with improved review quality when compared with traditional methods. However, reviewer-level interventions were also associated with increased duration of the peer-review process, with the only exception of revealing the identity of the reviewers. In individual studies, the only interventions found to have a significant effect on the peer-review process were the addition of a statistical reviewer, the use of appropriate checklists/guidelines, the editorial pre-screening of manuscripts, the assignment of a shorter deadline to accept the invitation to review, and the blinding of the reviewers to authors' identity (Table 2). No effect was demonstrated for all the other strategies.

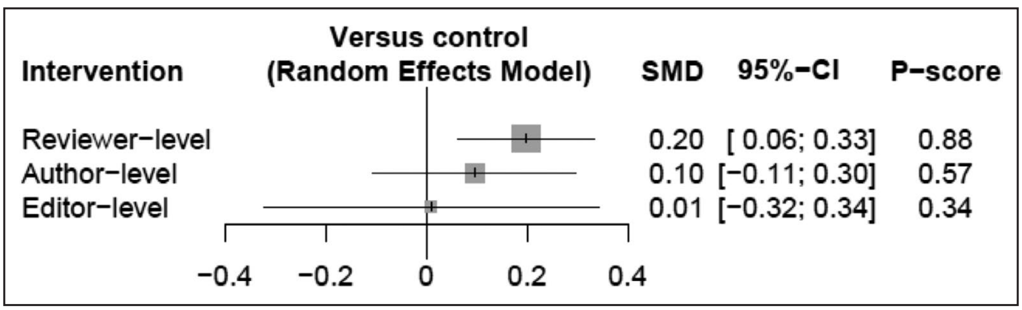


Figure 1. Network forest plot for quality of the peer-review process among the different interventions (random-effects model).
Larger *P* values signify larger standardized mean difference vs control and larger intervention effect on peer-review quality. SMD indicates standardized mean difference.

Table 3. League Table for the Outcomes of Peer-Review Quality and Length of Peer-Review Process

Peer-Review Quality			
Reviewer-level			
0.10 [−0.14 to 0.35]	Author-level		
0.19 [−0.17 to 0.55]	0.09 [−0.30 to 0.47]	Editor-level	
0.20 [0.06 to 0.33]	0.10 [−0.11 to 0.30]	0.01 [−0.32 to 0.34]	Control
Length of the Peer Review Process			
Reviewer-level			
−0.02 [−0.39 to 0.34]	Author-level		
−0.05 [−0.66 to 0.57]	−0.02 [−0.71 to 0.66]	Editor-level	
0.15 [0.01 to 0.29]	0.17 [−0.16 to 0.51]	0.19 [−0.40 to 0.79]	Control

Values in brackets represent 95% CI.

With almost 30 000 journals indexed in PubMed and scientific publication guiding medical practice, the importance of peer-review in medical journals cannot be underestimated.[30] However, only limited research on it has been published to date. In 2012 Larson and Chung[31] performed a systematic review of articles on peer-review of scientific manuscripts and found that out of 37 included papers, the great majority (78%) were editorials or commentaries that did not include original data.

In the only other systematic review and meta-analysis on the topic, Bruce et al found that the addition of a statistical reviewer and the use of open peer-review were associated with an increase in the quality of review.[5] Compared with their work, we have included 2 additional trials, grouped the intervention by their level in the process, and used a network meta-analysis to allow for direct and indirect comparisons and increase analytic power because of the relatively low number of available studies. It is concerning to note how, over the course of 3 decades, only 24 trials, mostly small, were performed to investigate a process that has immense implications for the medical community and the society at large. We believe that the most important finding of our analysis is that much more evidence is needed on such a crucial topic.

This is even more important as new concerns with regard to the integrity and quality of the peer-review process have recently emerged. A serious threat to good practice is represented by "predatory publishing", ie, an exploitive academic publishing business model based on journals that charge authors article processing fees and hijack the traditional peer-review processes by either manipulating peer-reviewer choice or fabricating reviews reports.[32] The dissatisfaction with the peer review system has led to an increasing use by authors of preprint servers, which however, raise concerns because of the absence of evaluation or certification of the published work (with the risk of unverified information being disseminated).[33]

A key issue rests with open review process, ie, the disclosure of reviewers' identity. While this approach may increase transparency and accountability, it may undermine the objectivity and thoroughness of reviewers, especially junior ones without tenure appointments. Also, during the current COVID-19 pandemic the traditional mechanisms of control that major scientific journals use have been stressed to their limits, and have sometimes failed.[34] Indeed, there is a clear conflict between the need to timely revise and possibly publish manuscripts

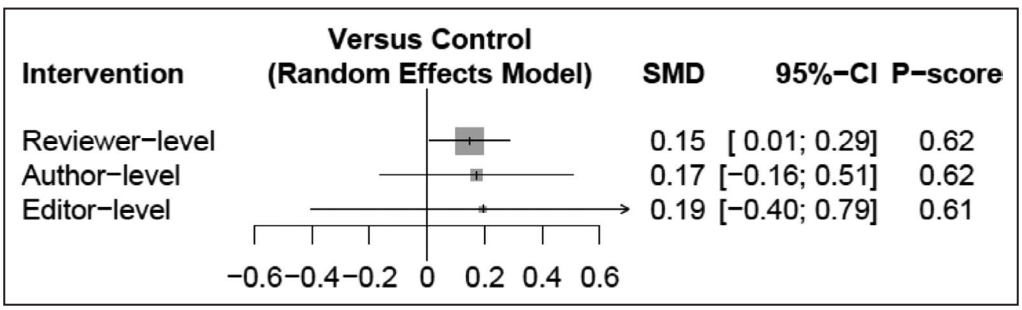


Figure 2. Network forest plot for peer-review duration among the different levels of interventions (random-effects model).
Larger *P* values signify larger standardized mean difference vs control and larger intervention effect on peer-review duration. SMD indicates standardized mean difference.

and safeguarding a thorough and valid peer-review process.

Limitations

The present analysis has several limitations. First, this review, as any similar work, provides more accurate estimates of effect than each included primary study, but cannot generate additional insights. Furthermore, it must be noted that the concept of "quality" of peer-review process is subjective by definition. There were important differences in interventions, journals, publishing models, as well as medical fields and outcomes among the included trials. While attempts were made to standardize the outcome definitions, heterogeneity between the studies remained. Most importantly, review quality is not necessarily related to manuscript quality and clinical importance. Because the number of studies for the individual interventions is limited some of the comparisons are underpowered. Finally, no trial included had a specific cardiovascular focus, but it seems likely that their results can be effectively applied to cardiovascular peer-review.

CONCLUSIONS

Limited information is available on the efficacy of interventions aimed at improving the peer-review process. Actions at reviewer-, rather than author- or editor-level seem to be the most effective, but further investigation into this important area is crucially needed.

ARTICLE INFORMATION

Received October 22, 2020; accepted February 18, 2021.

Affiliations

Department of Cardiothoracic Surgery (M.G., N.B.R., A.D.F., I.H., A.N., L.N.G.); and Samuel J. Wood Library and C.V. Starr Biomedical Information Centre (M.D.), Weill Cornell Medicine, New York, NY; Department of Medical-Surgical Sciences and Biotechnologies, Sapienza University of Rome, Latina, Italy (G.F., G.B.); IRCCS NEUROMED, Pozzilli, Isernia, Italy (G.F.); Schulich Heart Centre, Sunnybrook Health Science University of Toronto, Toronto, Canada (S.E.F.); and Mediterranea Cardiocentro, Napoli, Italy (G.B.).

Acknowledgments

Author contributions: Gaudino and Biondi-Zoccai were responsible for study design, data analysis, writing, critical review, and gave final approval. Robinson, Di Franco, Hameed, Naik, Demetres, were responsible for study selection, data extraction, writing, critical review, and gave final approval. Girardi, Frati, Fremes participated in study design, data analysis, manuscript drafting, critical review, and gave final approval.

Sources of Funding

None.

Disclosures

None.

Supplementary Material

Tables S1–S14
Figures S1–S15

REFERENCES

1. Kelly J, Sadeghieh T, Adeli K. Peer review in scientific publications: benefits, critiques, & a survival guide. *EJIFCC*. 2014;25:227–243.
2. Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med*. 2006;99:178–182. DOI: 10.1177/014107680609900414.
3. Huisman J, Smits J. Duration and quality of the peer review process: the author's perspective. *Scientometrics*. 2017;113:633–650. DOI: 10.1007/s11192-017-2310-5.
4. Kovanis M, Trinquart L, Ravaud P, Porcher R. Evaluating alternative systems of peer review: a large-scale agent-based modelling approach to scientific publication. *Scientometrics*. 2017;113:651–671. DOI: 10.1007/s11192-017-2375-1.
5. Bruce R, Chauvin A, Trinquart L, Ravaud P, Boutron I. Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. *BMC Med*. 2016;14:85. DOI: 10.1186/s12916-016-0631-5.
6. Alam M, Kim NA, Havey J, Rademaker A, Ratner D, Tregre B, West DP, Coleman WP. Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: a randomized multi-rater study: blinded vs. unblinded peer review. *Br J Dermatol*. 2011;165:563–567. DOI: 10.1111/j.1365-2133.2011.10432.x.
7. Arnau C, Cobo E, Maria Ribera J, Cardellach F, Selva A, Urrutia A. Efecto de la revisión estadística en la calidad de los manuscritos publicados en MEDICINA CLÍNICA: estudio aleatorizado. *Med Clín (Barc)*. 2003;121:690–694. DOI: 10.1016/S0025-7753(03)74064-0.
8. Callaham ML, Schriger DL. Effect of structured workshop training on subsequent performance of journal peer reviewers. *Ann Emerg Med*. 2002;40:323–328. DOI: 10.1067/mem.2002.127121.
9. Callaham ML, Knopp RK, Gallagher EJ. Effect of written feedback by editors on quality of reviews: two randomized trials. *JAMA*. 2002;287:2781–2783. DOI: 10.1001/jama.287.21.2781.
10. Cobo E, Selva-O'Callagham A, Ribera J-M, Cardellach F, Dominguez R, Vilardell M. Statistical reviewers improve reporting in biomedical articles: a randomized trial. *PLoS One*. 2007;2:e332. DOI: 10.1371/journal.pone.0000332.
11. Cobo E, Cortes J, Ribera JM, Cardellach F, Selva-O'Callaghan A, Kostov B, Garcia L, Cirugeda L, Altman DG, Gonzalez JA, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomised trial. *BMJ*. 2011;343:d6783. DOI: 10.1136/bmj.d6783.
12. Das Sinha S, Sahni P, Nundy S. Does exchanging comments of Indian and non-Indian reviewers improve the quality of manuscript reviews? *Natl Med J India*. 1999;12:210–213.
13. Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. *JAMA*. 1994;272:143–146. DOI: 10.1001/jama.1994.03520020069019.
14. Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA*. 1998;280:237–240. DOI: 10.1001/jama.280.3.237.
15. Houry D, Green S, Callaham M. Does mentoring new peer reviewers improve review quality? A randomized trial. *BMC Med Educ*. 2012;12:83. DOI: 10.1186/1472-6920-12-83.
16. John LK, Loewenstein G, Marder A, Callaham ML. Effect of revealing authors' conflicts of interests in peer review: randomized controlled trial. *BMJ*. 2019;367:l5896. DOI: 10.1136/bmj.l5896.
17. Johnston SC, Lowenstein DH, Ferriero DM, Messing RO, Oksenberg JR, Hauser SL, Stewart AF. Early editorial manuscript screening versus obligate peer review: a randomized trial. *Ann Neurol*. 2007;61:A10–A12. DOI: 10.1002/ana.21150.
18. Justice AC, Cho MK, Winker MA, Berlin JA, Rennie D. Does masking author identity improve peer review quality? A randomized controlled trial. PEER Investigators. *JAMA*. 1998;280:240–242. DOI: 10.1001/jama.280.3.240.
19. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. A randomized trial. *JAMA*. 1990;263:1371–1376. DOI: 10.1001/jama.1990.03440100079012.
20. Neuhauser D, Koran CJ. Calling medical care reviewers first. A randomized trial. *Med Care*. 1989;27:664–666. DOI: 10.1097/00005650-198906000-00009.
21. Okike K, Hug KT, Kocher MS, Leopold SS. Single-blind vs double-blind peer review in the setting of author prestige. *JAMA*. 2016;316:1315. DOI: 10.1001/jama.2016.11014.

22. Pitkin RM, Burmeister LF. Identifying manuscript reviewers: randomized comparison of asking first or just sending. *JAMA*. 2002;287:2795–2796. DOI: 10.1001/jama.287.21.2795.
23. Provenzale JM. A shorter invitation period for AJR manuscript reviewers: impact on time to completion of reviews. *AJR Am J Roentgenol*. 2020;214:37–40. DOI: 10.2214/AJR.19.21358.
24. Schroter S, Black N, Evans S, Carpenter J, Godlee F, Smith R. Effects of training on quality of peer review: randomised controlled trial. *BMJ*. 2004;328:673. DOI: 10.1136/bmj.38023.700775.AE.
25. van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*. 1998;280:234–237. DOI: 10.1001/jama.280.3.234.
26. van Rooyen S, Godlee F, Evans S, Black N, Smith R. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ*. 1999;318:23–27. DOI: 10.1136/bmj.318.7175.23.
27. van Rooyen S, Delamothe T, Evans SJW. Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial. *BMJ*. 2010;341:c5729. DOI: 10.1136/bmj.c5729.
28. Vinther S, Nielsen OH, Rosenberg J, Keiding N, Schroeder TV. Same review quality in open versus blinded peer review in "Ugeskrift for Læger". *Dan Med J*. 2012;59:A4479.
29. Walsh E, Rooney M, Appleby L, Wilkinson G. Open peer review: a randomised controlled trial. *Br J Psychiatry*. 2000;176:47–51. DOI: 10.1192/bjp.176.1.47.
30. List of All Journals Cited in PubMed®. Available at: https://www.nlm.nih.gov/bsd/serfile_addedinfo.html. Accessed April 22, 2020.
31. Larson BP, Chung KC. A systematic review of peer review for scientific manuscripts. *Hand (N Y)*. 2012;7:37–44. DOI: 10.1007/s11552-012-9392-6.
32. Ferris LE, Winker MA. Ethical issues in publishing in predatory journals. *Biochem Med*. 2017;27:279–284. DOI: 10.11613/BM.2017.030.
33. Preprint servers: challenges and consequences. INLEXIO. Published January 23, 2017. Available at: https://www.inlexio.com/preprint-servers-challenges-consequences/. Accessed June 6, 2020.
34. Gumbrecht J, Fox M. Two coronavirus studies retracted after questions emerge about data. CNN. Available at: https://www.cnn.com/2020/06/04/health/retraction-coronavirus-studies-lancet-nejm/index.html. Accessed June 11, 2020.

SUPPLEMENTAL MATERIAL

Table S1. Full search strategy.

Ovid MEDLINE – ALL (1946 to June 12th, 2020)

Searched on June 12th, 2020

RCT Filter: BMJ Publishing Group Limited. BMJ Best Practice Study design search filters 2017 Available from: https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/study-design-search-filters/

1	*"Peer Review"/ or *"Peer Review, Research"/
2	(peer adj3 (review or reviewed or reviewing or reviewer or reviewers)).ti.
3	(blind review or blind reviewed or referee* or post-publication review or cascading review or third party review or author suggested reviewers or editor suggested reviewers or manuscript reviewer*).ti.
4	or/1-3
5	"randomized controlled trial".pt.
6	(random$ or placebo$ or single blind$ or double blind$ or triple blind$).ti,ab.
7	(retraction of publication or retracted publication).pt.
8	or/5-7
9	(animals not humans).sh.
10	((comment or editorial or meta-analysis or practice-guideline or review or letter) not "randomized controlled trial").pt.
11	(random sampl$ or random digit$ or random effect$ or random survey or random regression).ti,ab. not "randomized controlled trial".pt.
12	8 not (9 or 10 or 11)
13	4 and 12

Table S2. **The Cochrane Collaboration's tool for assessing risk of bias in randomized trials.**

	RANDOM SEQUENCE GENERATION	ALLOCATION CONCEALMENT	BLINDING OF PARTICIPANTS	BLINDING OF OUTCOME ASSESSMENT	INCOMPLETE OUTCOME DATA	SELECTIVE REPORTING	OTHER SOURCES OF BIAS
Alam, 2011	+	?	-	+	+	?	?
Arnau, 2003	+	+	+	+	+	?	?
Callaham, 2002	+	?	+	+	+	?	?
Callaham, 2002 JAMA	+	?	+	+	+	?	?
Cobo, 2011	+	+	+	+	+	+	?
Cobo, 2007	+	?	+	+	+	?	?
Das Sinha, 1999	+	?	-	+	+	?	?
Fisher, 1994	+	?	-	+	+	?	?
Godlee, 1998	+	?	-	+	+	?	?
Houry, 2012	+	?	?	+	+	?	?
John, 2019	+	?	+	+	+	?	?
Johnston, 2007	+	?	+	+	+	?	?
Justice, 1998	+	?	-	+	+	?	?
McNutt, 1990	+	?	-	+	+	?	?
Neuhauser, 1989	+	?	?	?	+	?	?
Okike, 2016	+	?	-	?	+	+	?
Pitkin, 2002	+	?	?	+	+	?	?
Provenzale, 2020	?	?	?	+	+	?	?
Schroter, 2004	+	?	-	+	+	?	?
Van Rooyen, 2010	+	?	+	+	+	?	?
Van Rooyen, 1998	+	?	-	+	+	?	?
Van Rooyen, 1999	+	?	?	+	+	?	?
Vinther, 2012	+	?	?	+	+	?	?
Walsh, 2000	+	?	?	+	+	?	?

+	**Low Risk**
?	**Uncertain**
-	**High Risk**

Table S3. Netsplit for the different peer-review interventions for the main outcome of peer-review quality (random effects model). Example: 19 studies compared standard peer-review process vs reviewer-level interventions and the estimated treatment effect (standardized mean difference [SMD]) was -0.198.

Random effects model:

comparison	k	prop	nma	direct	indir.	Diff	z	p-value
Author-level:Control	9	1.00	0.0958	0.0958	.	.	.	.
Author-level:Editor-level	0	0	0.0862	.	0.0862	.	.	.
Author-level:Reviewer-level	0	0	-0.1022	.	-0.1022	.	.	.
Control:Editor-level	5	1.00	-0.0096	-0.0096	.	.	.	.
Control:Reviewer-level	19	1.00	-0.1980	-0.1980	.	.	.	.
Editor-level:Reviewer-level	0	0	-0.1884	.	-0.1884	.	.	.

Legend:
- comparison - Treatment comparison
- k - Number of studies providing direct evidence
- prop - Direct evidence proportion
- nma - Estimated treatment effect (SMD) in network meta-analysis
- direct - Estimated treatment effect (SMD) derived from direct evidence
- indir. - Estimated treatment effect (SMD) derived from indirect evidence
- Diff - Difference between direct and indirect treatment estimates
- z - z-value of test for disagreement (direct versus indirect)
- p-value - p-value of test for disagreement (direct versus indirect)

Table S4. Quantifying heterogeneity/inconsistency, tests of heterogeneity (within designs) and inconsistency (between designs), and design-specific decomposition of within-designs Q statistic (main outcome of peer-review quality).

```
                     Q df  p-value
Total           130.42 30 < 0.0001
Within designs  130.42 30 < 0.0001
Between designs   0.00  0       --

Design-specific decomposition of within-designs Q statistic

                 Design      Q df  p-value
   Author-level:Control  17.57  8   0.0247
   Control:Editor-level   9.18  4   0.0568
 Control:Reviewer-level 103.68 18 < 0.0001

Q statistic to assess consistency under the assumption of
a full design-by-treatment interaction random effects model

                   Q df p-value tau.within tau2.within
Between designs 0.00  0      --     0.2525      0.0638
```

Table S5. League table for the main outcome of peer-review quality (fixed effect model).

Reviewer-level			
0.07 [-0.04; 0.17]	Author-level		
0.11 [-0.13; 0.36]	0.05 [-0.21; 0.30]	Editor-level	
0.14 [0.08; 0.20]	0.07 [-0.01; 0.16]	0.03 [-0.21; 0.27]	Control

Table S6. Netsplit for the different peer-review interventions for the main outcome of peer-review quality (fixed effect model). Example: 19 studies compared standard peer-review process vs reviewer-level interventions and the estimated treatment effect (standardized mean difference [SMD]) was -0.1397.

Fixed effect model:

comparison	k	prop	nma	direct	indir.	Diff	z	p-value
Author-level:Control	9	1.00	0.0738	0.0738	.	.	.	.
Author-level:Editor-level	0	0	0.0460	.	0.0460	.	.	.
Author-level:Reviewer-level	0	0	-0.0659	.	-0.0659	.	.	.
Control:Editor-level	5	1.00	-0.0278	-0.0278	.	.	.	.
Control:Reviewer-level	19	1.00	-0.1397	-0.1397	.	.	.	.
Editor-level:Reviewer-level	0	0	-0.1120	.	-0.1120	.	.	.

```
Legend:
 comparison - Treatment comparison
 k          - Number of studies providing direct evidence
 prop       - Direct evidence proportion
 nma        - Estimated treatment effect (SMD) in network meta-analysis
 direct     - Estimated treatment effect (SMD) derived from direct evidence
 indir.     - Estimated treatment effect (SMD) derived from indirect evidence
 Diff       - Difference between direct and indirect treatment estimates
 z          - z-value of test for disagreement (direct versus indirect)
 p-value    - p-value of test for disagreement (direct versus indirect)
```

Table S7. League table for the duration of the peer-review process (random effects model).

Reviewer-level			
-0.02 [-0.39; 0.34]	Author-level		
-0.05 [-0.66; 0.57]	-0.02 [-0.71; 0.66]	Editor-level	
0.15 [0.01; 0.29]	0.17 [-0.16; 0.51]	0.19 [-0.40; 0.79]	Control

Table S8. Netsplit for the different peer-review interventions for peer-review duration (random effects model). Example: 12 studies compared standard peer-review process vs reviewer-level interventions and the estimated treatment effect (standardized mean difference [SMD]) was -0.1481.

Random effects model:

comparison	k	prop	nma	direct	indir.	Diff	z	p-value
Author-level:Control	2	1.00	0.1710	0.1710	.	.	.	.
Author-level:Editor-level	0	0	-0.0234	.	-0.0234	.	.	.
Author-level:Reviewer-level	0	0	0.0228	.	0.0228	.	.	.
Control:Editor-level	1	1.00	-0.1943	-0.1943	.	.	.	.
Control:Reviewer-level	12	1.00	-0.1481	-0.1481	.	.	.	.
Editor-level:Reviewer-level	0	0	0.0462	.	0.0462	.	.	.

```
Legend:
 comparison - Treatment comparison
 k          - Number of studies providing direct evidence
 prop       - Direct evidence proportion
 nma        - Estimated treatment effect (SMD) in network meta-analysis
 direct     - Estimated treatment effect (SMD) derived from direct evidence
 indir.     - Estimated treatment effect (SMD) derived from indirect evidence
 Diff       - Difference between direct and indirect treatment estimates
 z          - z-value of test for disagreement (direct versus indirect)
 p-value    - p-value of test for disagreement (direct versus indirect)
```

Table S9. Quantifying heterogeneity/inconsistency, tests of heterogeneity (within designs) and inconsistency (between designs), and design-specific decomposition of within-designs Q statistic (peer-review duration).

```
Q statistics to assess homogeneity / consistency

                    Q df  p-value
Total           60.62 12 < 0.0001
Within designs  60.62 12 < 0.0001
Between designs  0.00  0       --

Design-specific decomposition of within-designs Q statistic

                 Design     Q df  p-value
   Author-level:Control  4.46  1   0.0347
 Control:Reviewer-level 56.16 11 < 0.0001

Q statistic to assess consistency under the assumption of
a full design-by-treatment interaction random effects model

                    Q df p-value tau.within tau2.within
Between designs -0.00  0      --     0.2184      0.0477
```

Table S10. League table for the duration of the peer-review process (fixed effect model).

Reviewer-level			
0.01 [-0.13; 0.16]	Author-level		
-0.07 [-0.49; 0.36]	-0.08 [-0.52; 0.36]	Editor-level	
0.13 [0.07; 0.19]	0.12 [-0.02; 0.25]	0.19 [-0.22; 0.61]	Control

Table S11. Netsplit for the different peer-review interventions for peer-review duration (fixed effect model). Example: 12 studies compared standard peer-review process vs reviewer-level interventions and the estimated treatment effect (standardized mean difference [SMD]) was -0.1293.

```
Fixed effect model:

                 comparison  k prop     nma  direct  indir. Diff z p-value
        Author-level:Control  2 1.00  0.1162  0.1162       .    . .       .
   Author-level:Editor-level  0    0 -0.0781       . -0.0781    . .       .
 Author-level:Reviewer-level  0    0 -0.0131       . -0.0131    . .       .
        Control:Editor-level  1 1.00 -0.1943 -0.1943       .    . .       .
      Control:Reviewer-level 12 1.00 -0.1293 -0.1293       .    . .       .
 Editor-level:Reviewer-level  0    0  0.0650       .  0.0650    . .       .
```

```
Legend:
 comparison - Treatment comparison
 k          - Number of studies providing direct evidence
 prop       - Direct evidence proportion
 nma        - Estimated treatment effect (SMD) in network meta-analysis
 direct     - Estimated treatment effect (SMD) derived from direct evidence
 indir.     - Estimated treatment effect (SMD) derived from indirect evidence
 Diff       - Difference between direct and indirect treatment estimates
 z          - z-value of test for disagreement (direct versus indirect)
 p-value    - p-value of test for disagreement (direct versus indirect)
```

Table S12. League table for the duration of the peer-review process (reviewer-level interventions; random effects model).

Blinding			
-0.28 [-0.62; 0.06]	Training		
-0.44 [-0.74; -0.15]	-0.17 [-0.54; 0.20]	Other_intervention	
0.01 [-0.17; 0.19]	0.28 [0.00; 0.57]	0.45 [0.21; 0.68]	Control

Table S13. Netsplit for the different peer-review interventions for peer-review duration (reviewer-level interventions, random effects model). Example: 8 studies compared blinding vs standard peer-review process and the estimated treatment effect (standardized mean difference [SMD]) was 0.0069.

Random effects model:

comparison	k	prop	nma	direct	indir.	Diff	z	p-value
Blinding:Control	8	1.00	0.0069	0.0069	.	.	.	.
Blinding:Other_intervention	0	0	-0.4422	.	-0.4422	.	.	.
Blinding:Training	0	0	-0.2767	.	-0.2767	.	.	.
Control:Other_intervention	6	1.00	-0.4491	-0.4491	.	.	.	.
Control:Training	5	1.00	-0.2836	-0.2836	.	.	.	.
Other_intervention:Training	0	0	0.1654	.	0.1654	.	.	.

```
Legend:
 comparison - Treatment comparison
 k          - Number of studies providing direct evidence
 prop       - Direct evidence proportion
 nma        - Estimated treatment effect (SMD) in network meta-analysis
 direct     - Estimated treatment effect (SMD) derived from direct evidence
 indir.     - Estimated treatment effect (SMD) derived from indirect evidence
 Diff       - Difference between direct and indirect treatment estimates
 z          - z-value of test for disagreement (direct versus indirect)
 p-value    - p-value of test for disagreement (direct versus indirect)
```

Table S14. Quantifying heterogeneity/inconsistency, tests of heterogeneity (within designs) and inconsistency (between designs), and design-specific decomposition of within-designs Q statistic (peer-review duration – reviewer-level interventions).

```
Q statistics to assess homogeneity / consistency

                    Q df  p-value
Total           67.72 16 < 0.0001
Within designs  67.72 16 < 0.0001
Between designs -0.00  0       --

Design-specific decomposition of within-designs Q statistic

                     Design     Q df  p-value
           Blinding:Control 14.23  7   0.0472
 Control:Other_intervention 51.20  5 < 0.0001
           Control:Training  2.29  4   0.6827

Q statistic to assess consistency under the assumption of
a full design-by-treatment interaction random effects model

                   Q df p-value tau.within tau2.within
Between designs 0.00  0      --     0.2338      0.0547
```

Figure S1. **Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart of our analysis.**

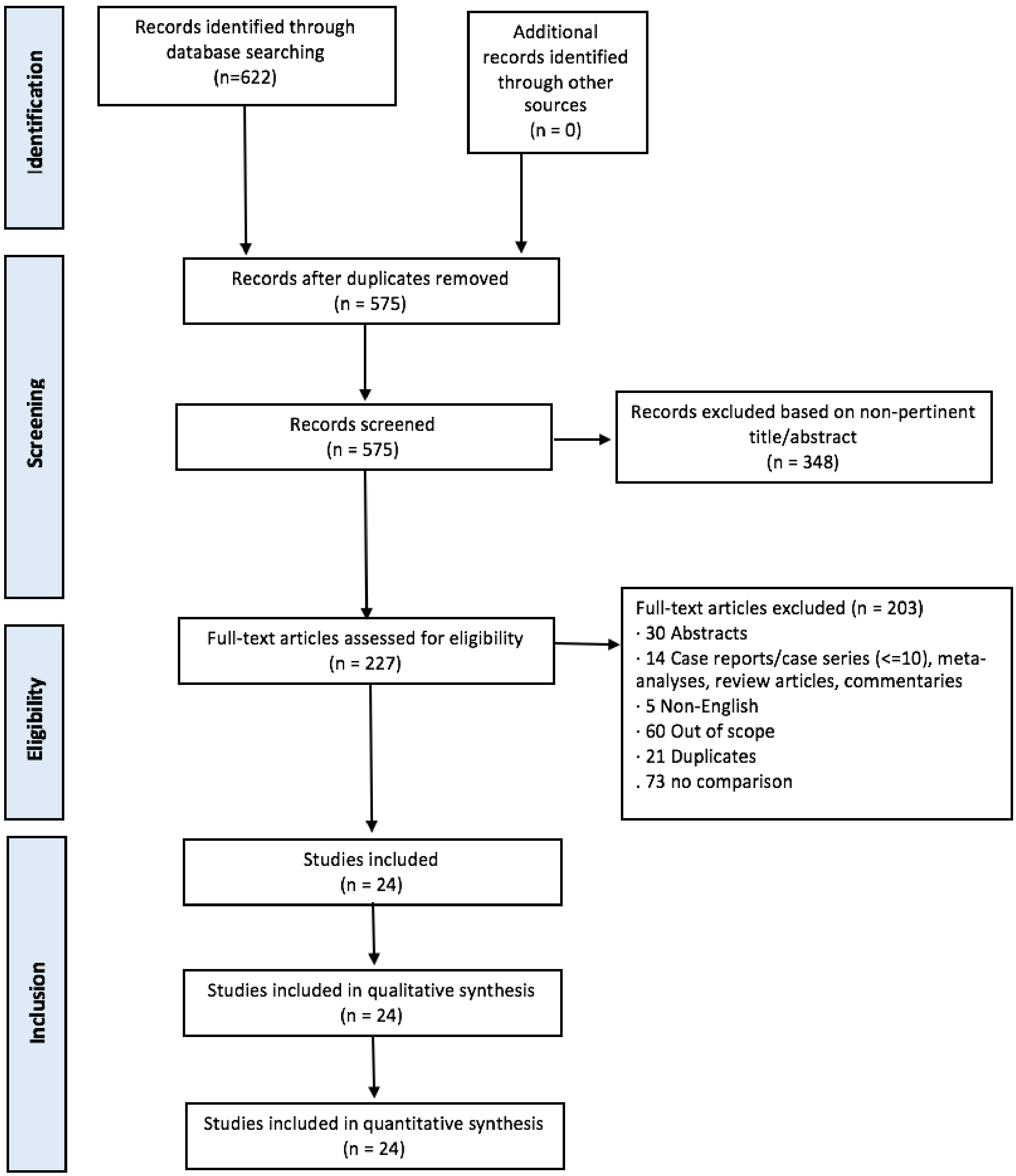


Figure S2. **Net graph for the main outcome of peer-review quality.**

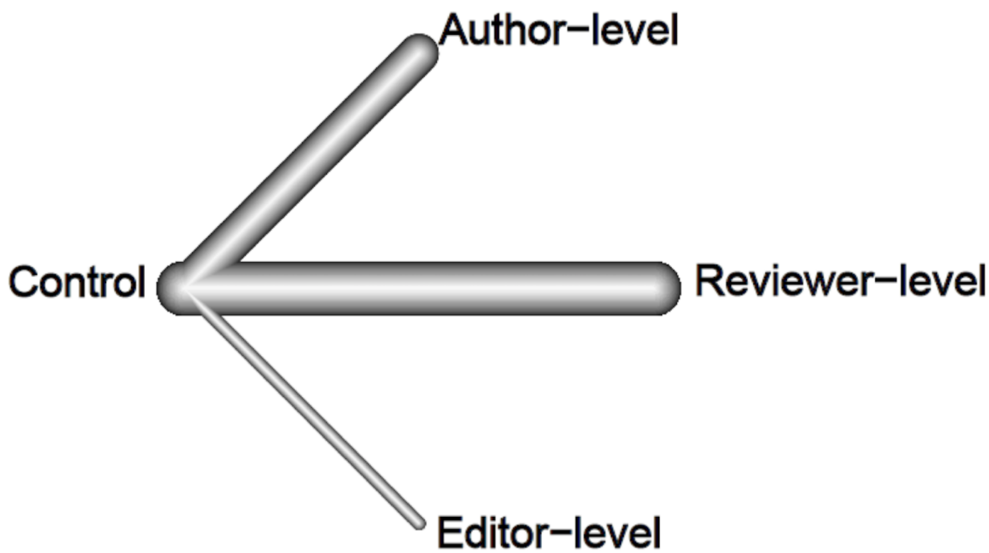


Figure S3. Funnel plot for the assessment of publication bias for the main outcome of peer-review quality.

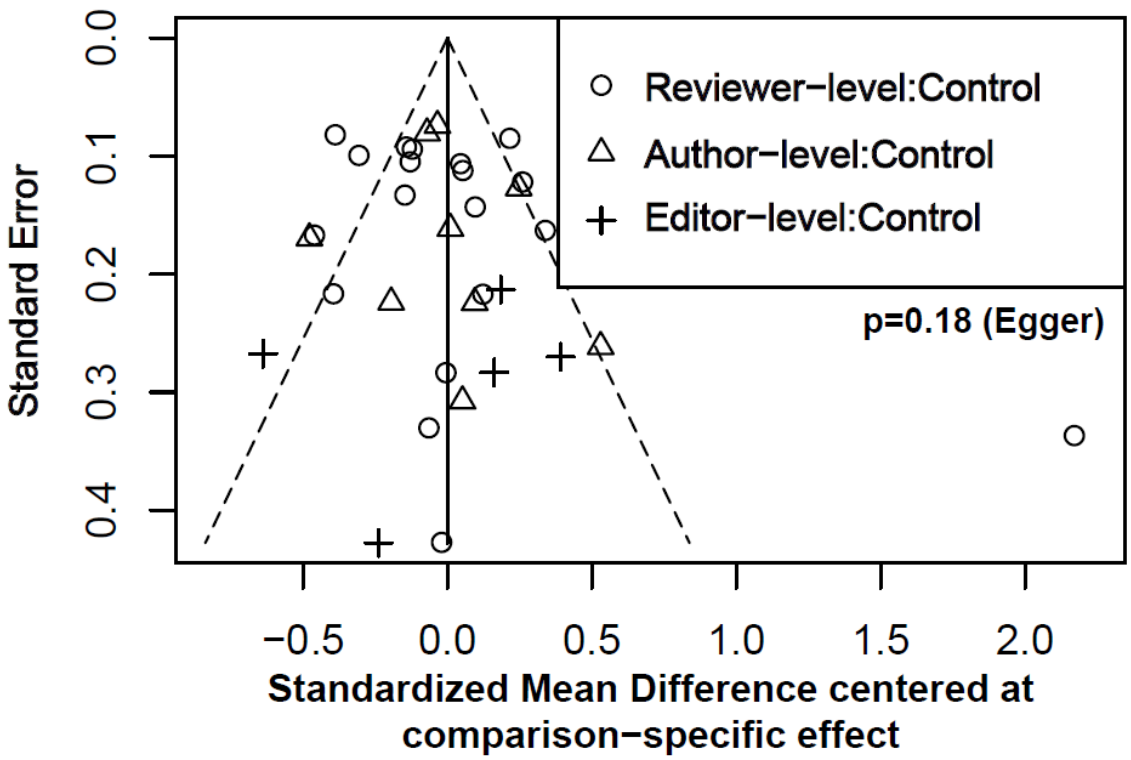
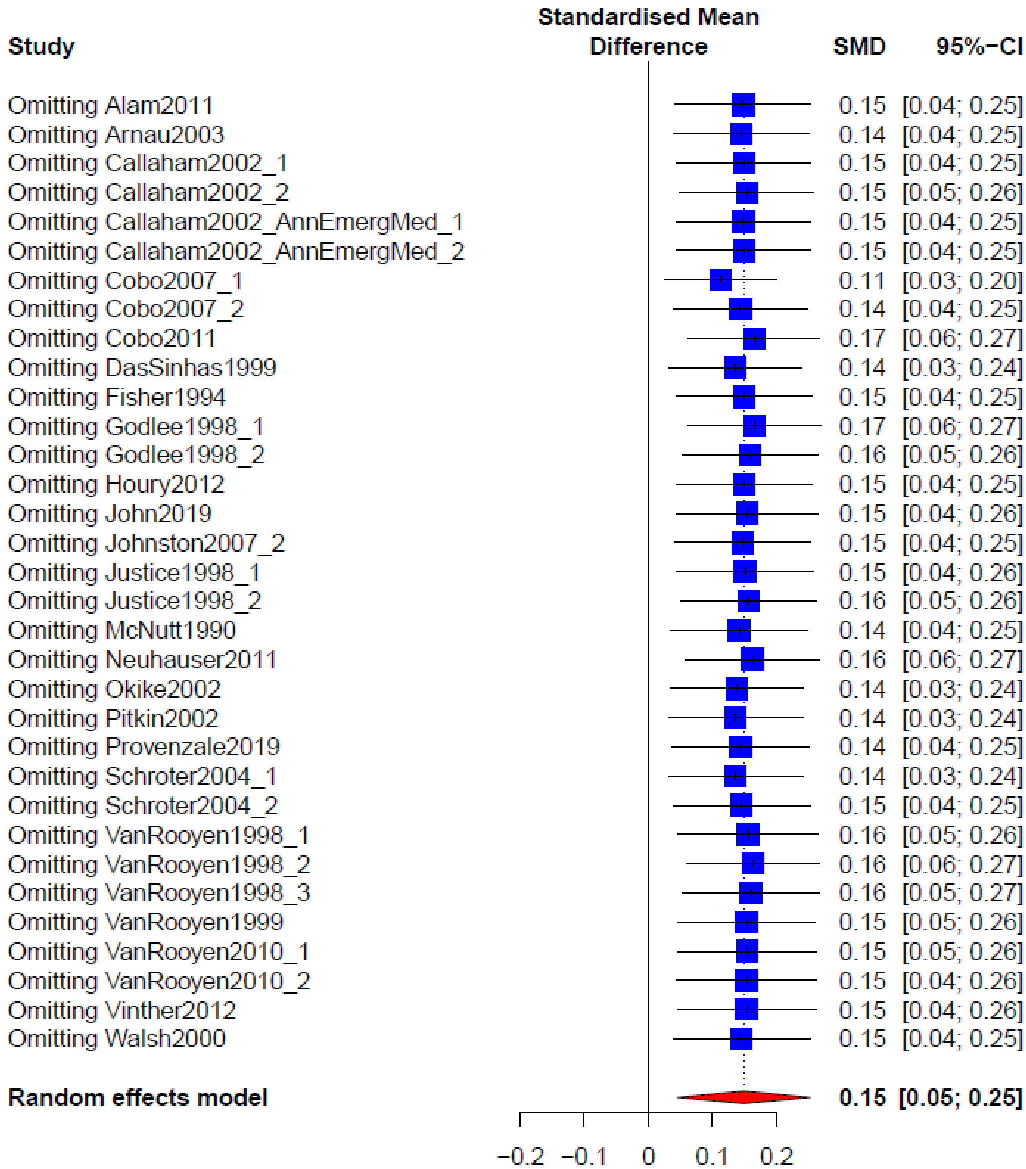
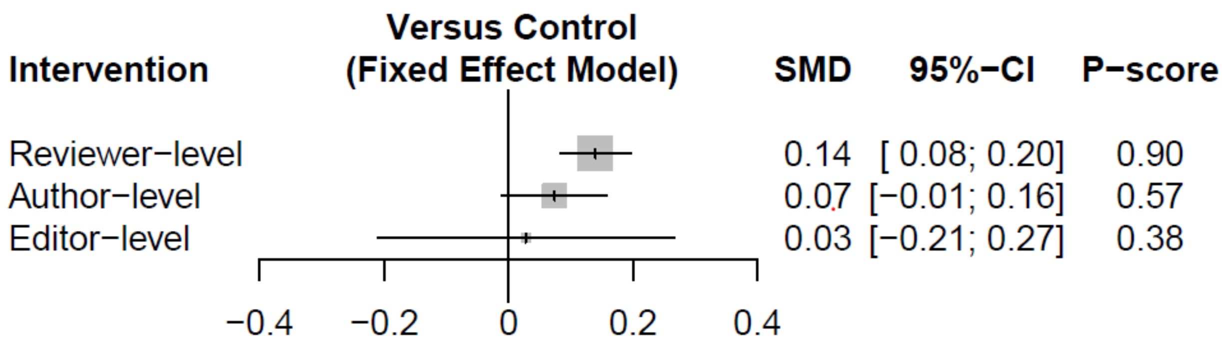


Figure S4. Leave-one-out analysis for standardized mean difference for the main outcome of peer-review quality (random effects model).



.

Figure S5. Network forest plot for quality of the peer-review process among the different interventions (fixed effect model).



Larger p-scores signify larger standardized mean difference (SMD) vs control and larger intervention effect on peer-review quality.

Figure S6. Leave-one-out analysis for standardized mean difference for the main outcome of peer-review quality (fixed effect model).

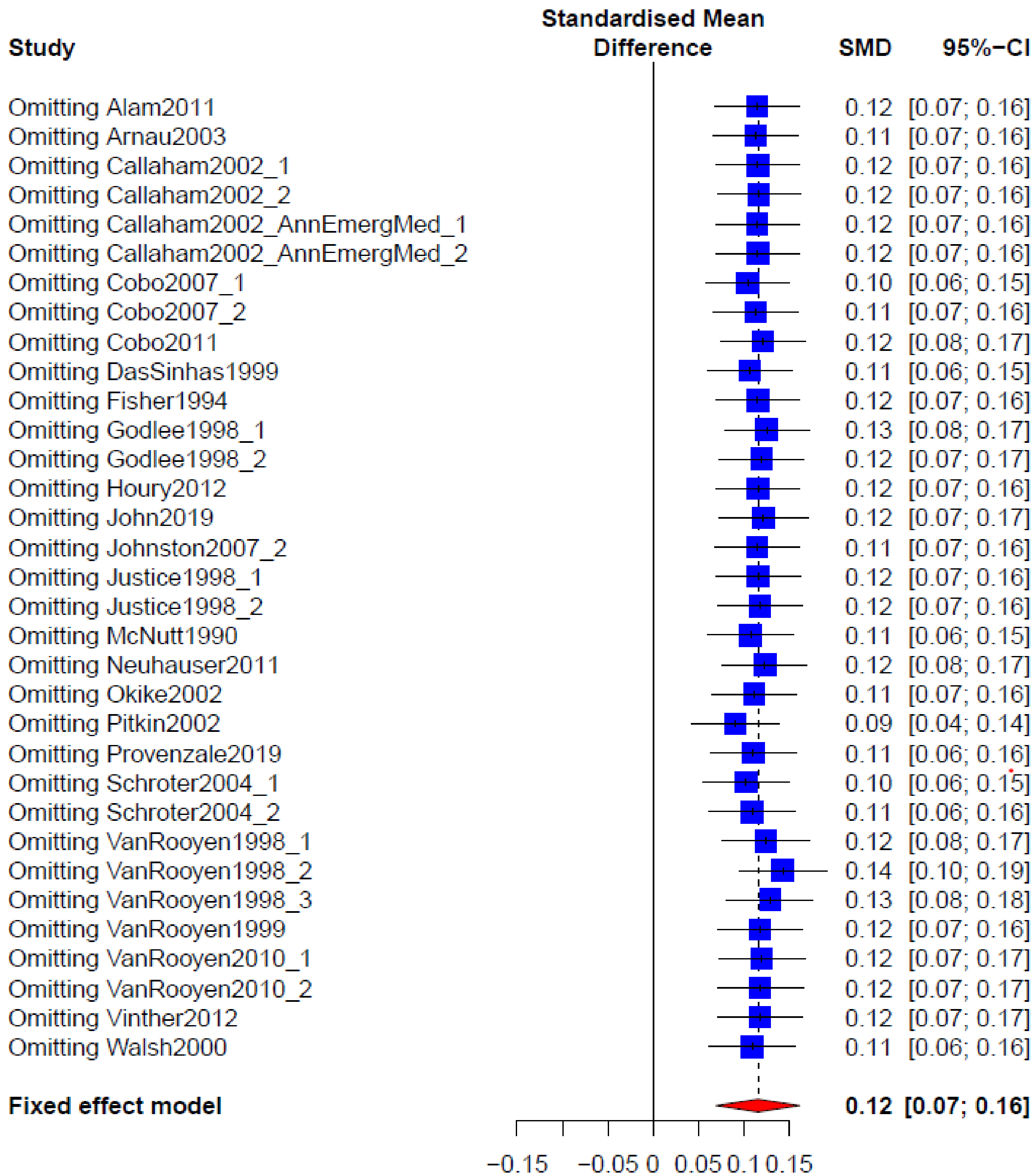


Figure S7. **Net graph for the duration of the peer-review process.**

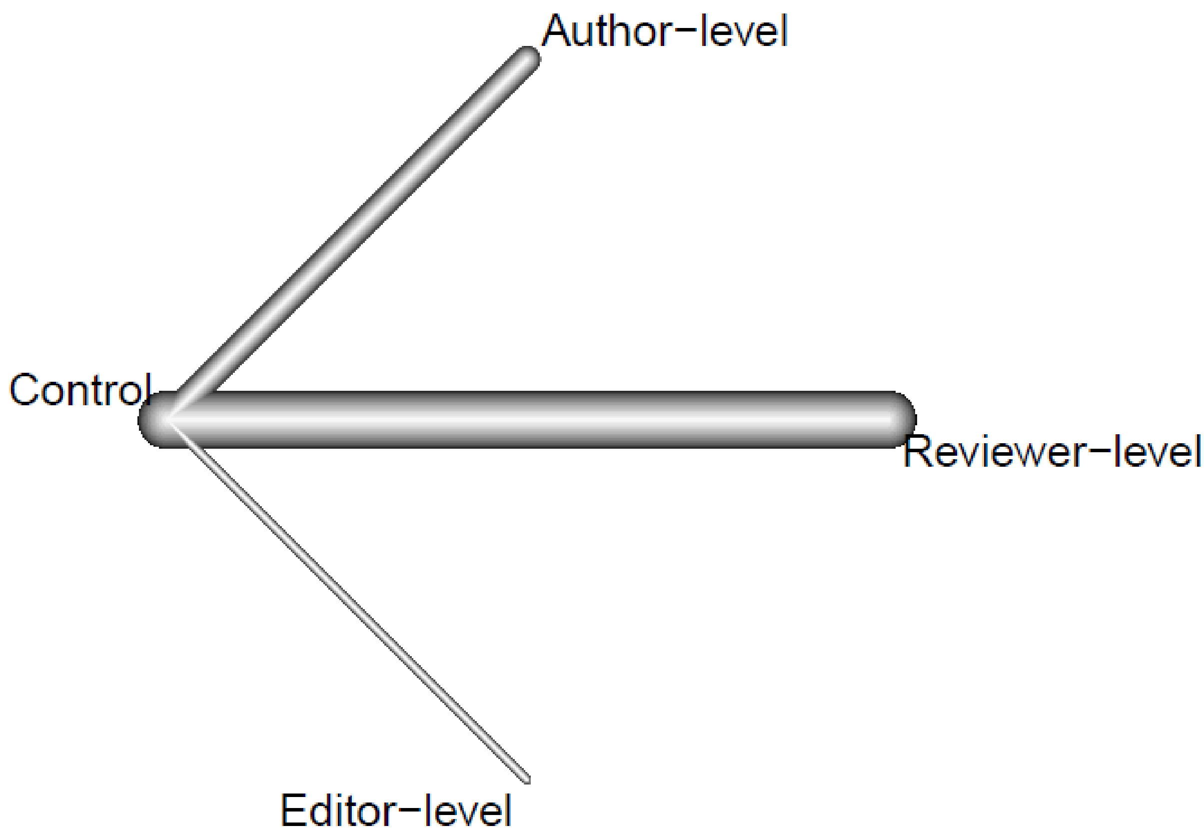


Figure S8. Funnel plot for the assessment of publication bias for duration of the peer-review process.

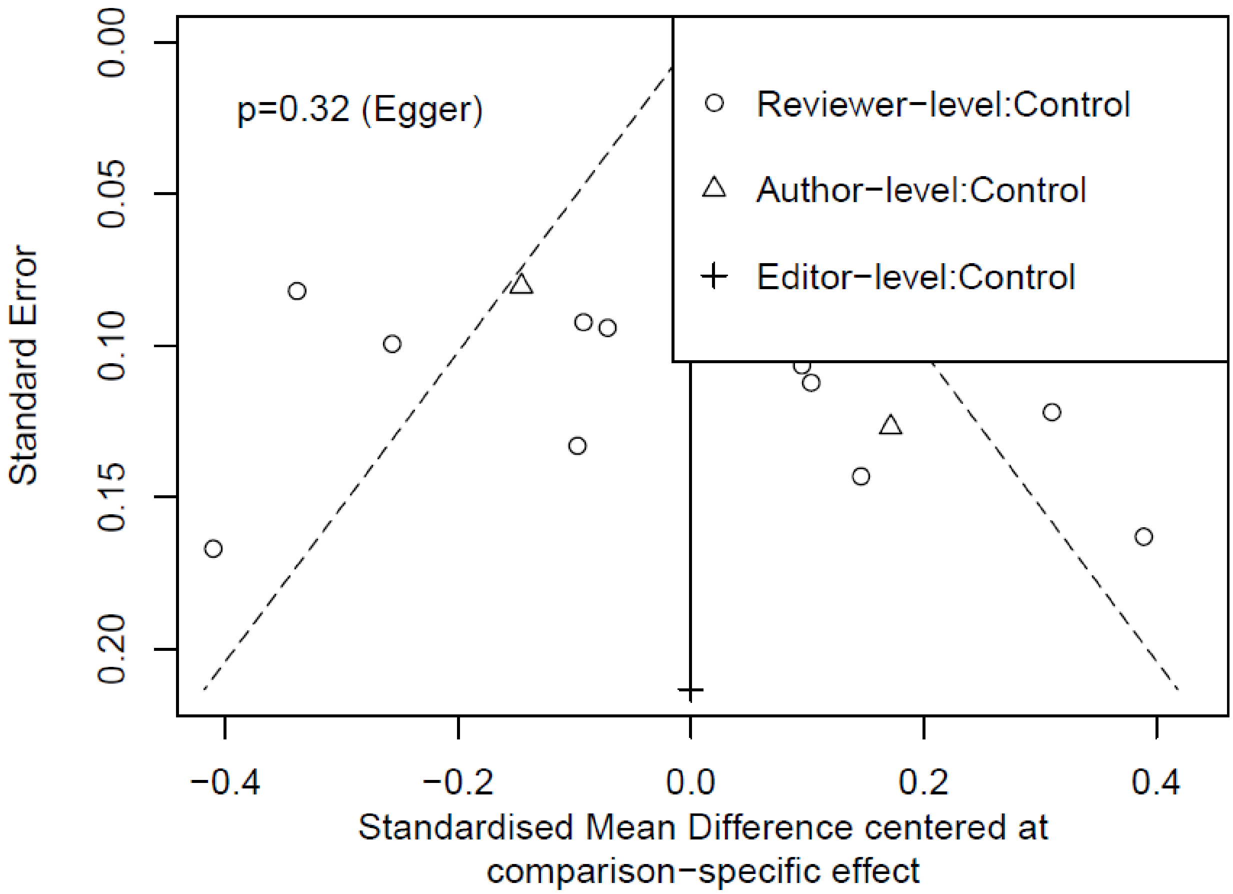


Figure S9. Leave-one-out analysis for standardized mean difference for peer-review duration (random effects model).

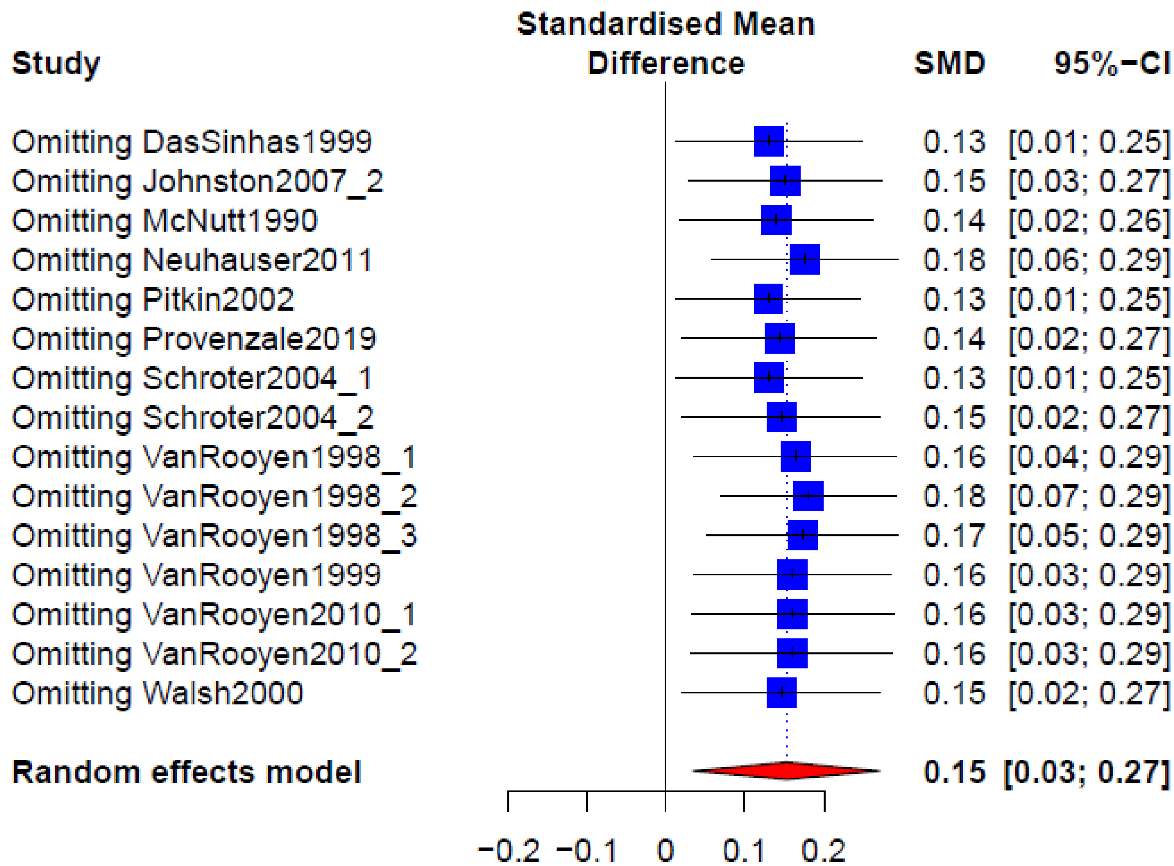
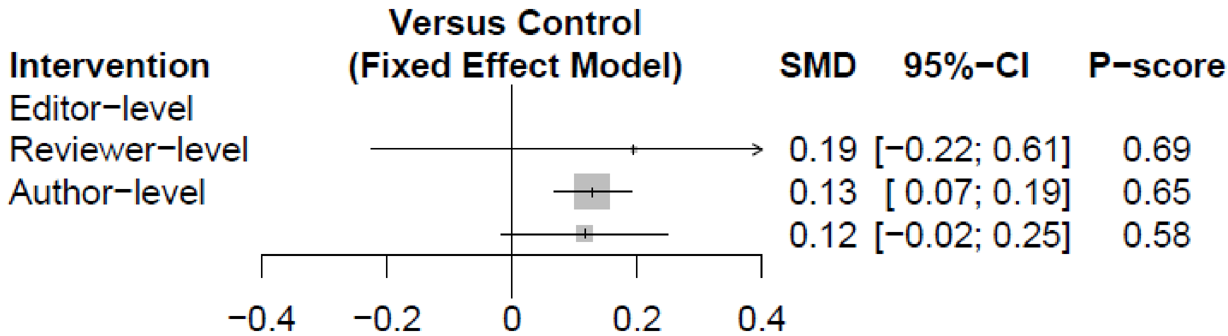


Figure S10. Network forest plot for peer-review duration among the different interventions (fixed effect model).



Larger p-scores signify larger standardized mean difference (SMD) vs control and larger intervention effect on peer-review quality.

Figure S11. Leave-one-out analysis for standardized mean difference for peer-review duration (fixed effect model).

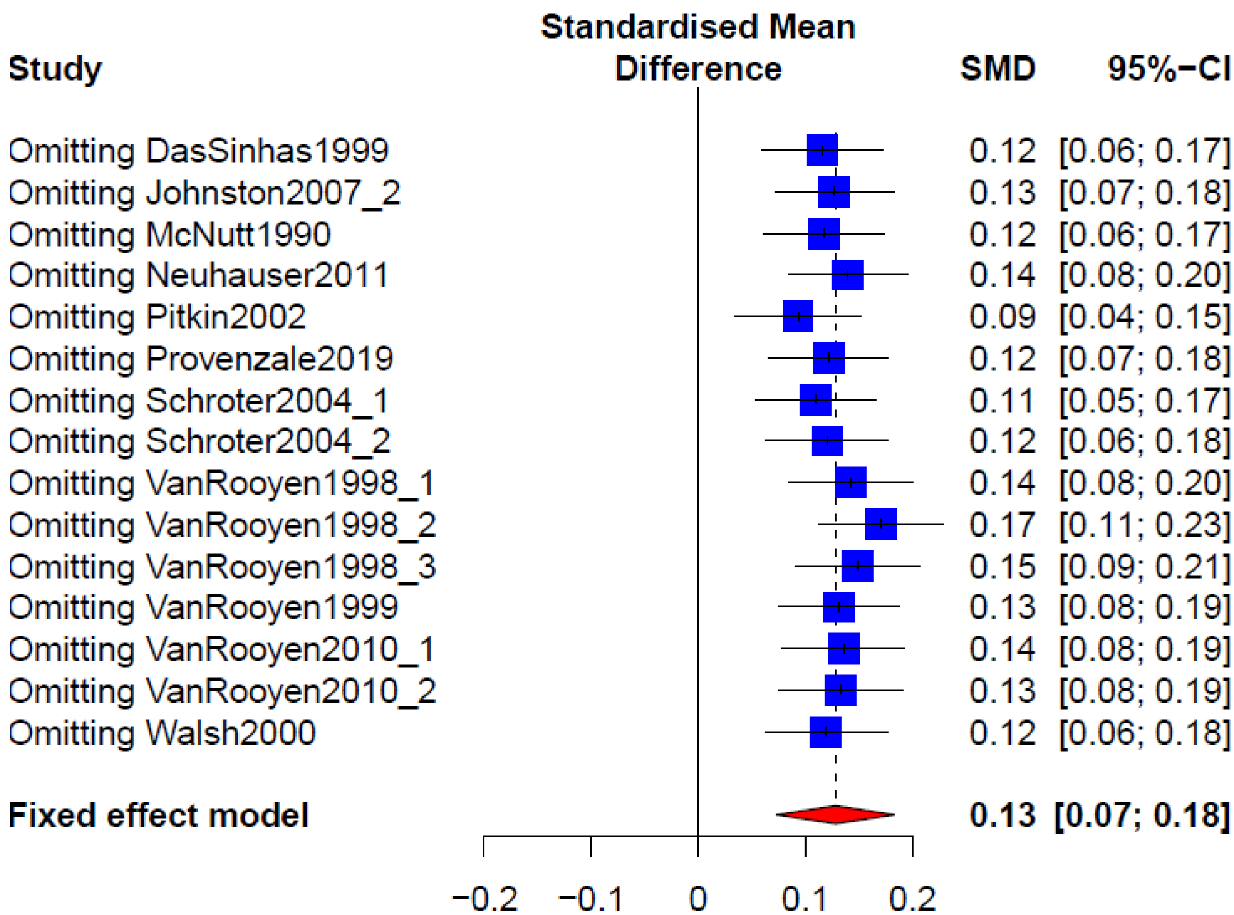
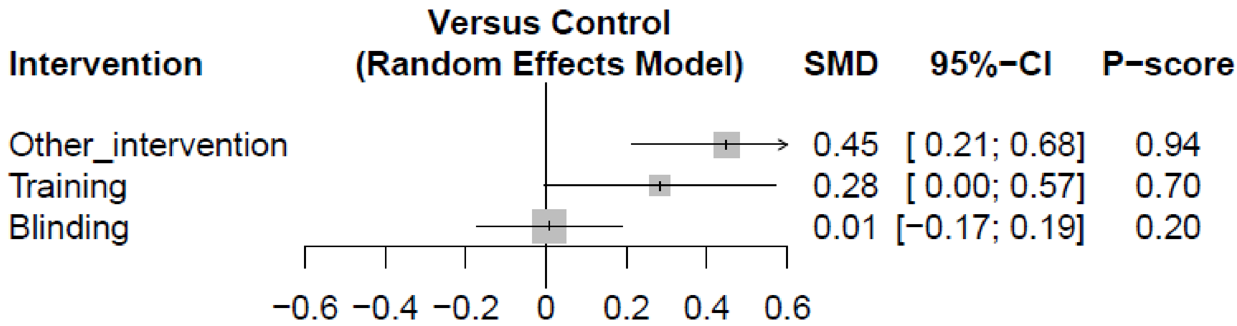


Figure S12. **Network forest plot for peer-review duration among the different reviewer-level interventions (random effects model).**



Larger p-scores signify larger standardized mean difference (SMD) vs control and larger intervention effect on peer-review duration.

Figure S13. **Net graph for the duration of the peer-review process (reviewer-level interventions).**

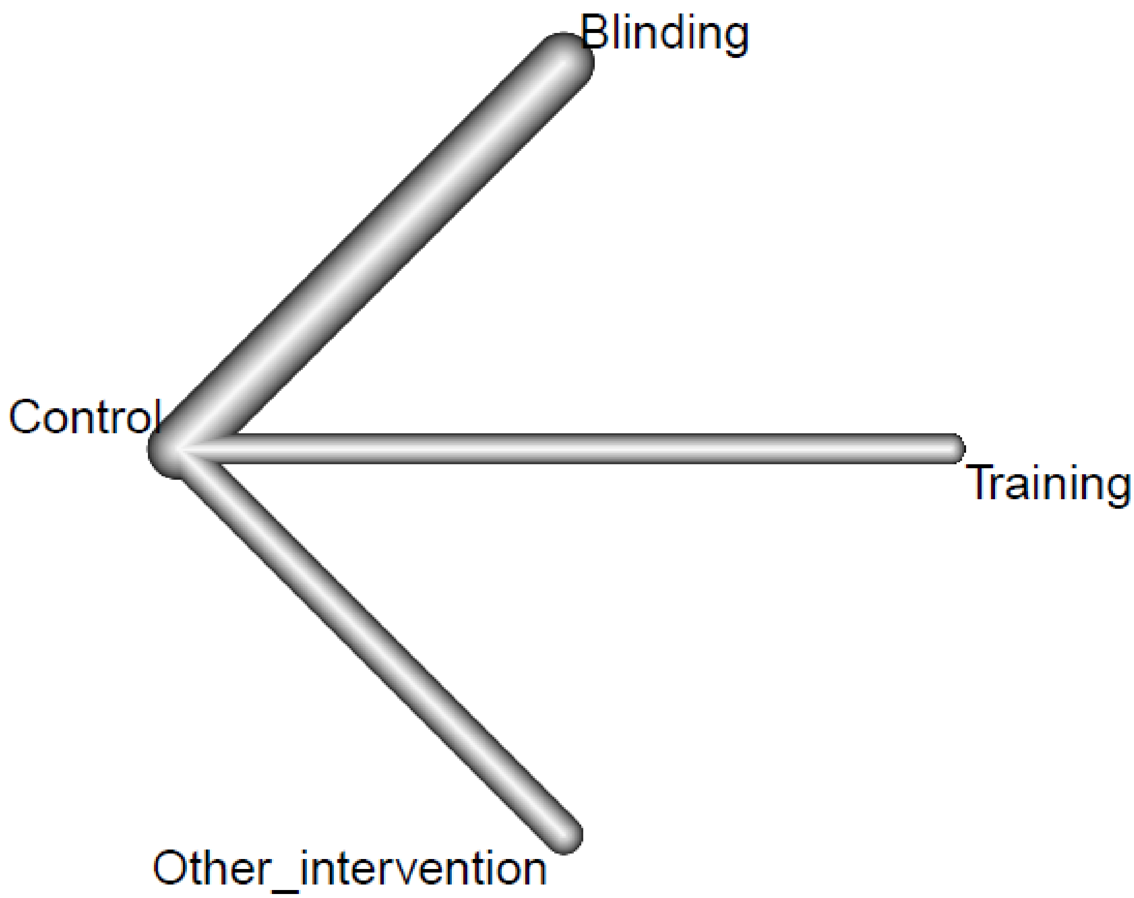


Figure S14. Funnel plot for the assessment of publication bias for duration of the peer-review process (reviewer-level interventions).

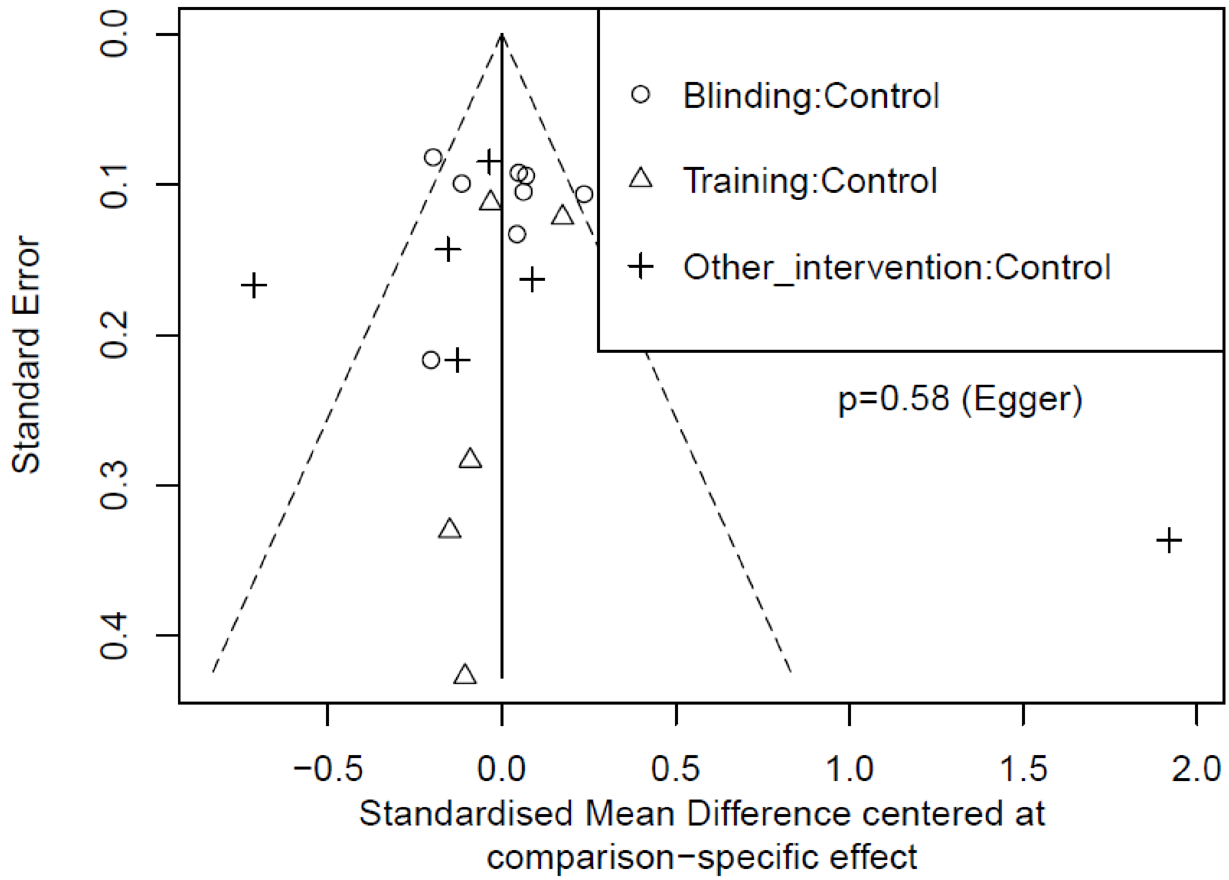


Figure S15. Leave-one-out analysis for standardized mean difference for peer-review duration (reviewer-level interventions; random effects model).