Cochrane Library

Cochrane Database of Systematic Reviews

Reviewer training for improving grant and journal peer review (Protocol)

Hesselberg JO, Dalsbø TK, Stromme H, Svege I, Fretheim A

Hesselberg J-O, Dalsbø TK, Stromme H, Svege I, Fretheim A.
Reviewer training for improving grant and journal peer review (Protocol).
Cochrane Database of Systematic Reviews 2020, Issue 11. Art. No.: MR000056.
DOI: 10.1002/14651858.MR000056.

www.cochranelibrary.com

TABLE OF CONTENTS

[Methodology Protocol]

Reviewer training for improving grant and journal peer review

Jan-Ole Hesselberg[1], Therese K Dalsbø[2], Hilde Stromme[3], Ida Svege[4], Atle Fretheim[5]

[1]Department of Psychology, University of Oslo, Oslo, Norway. [2]Norwegian Knowledge Centre for the Health Services, Oslo, Norway. [3]Medical Library, University of Oslo, Oslo, Norway. [4]Stiftelsen Dam, Oslo, Norway. [5]Institute of Health and Society, University of Oslo, Oslo, Norway

Contact address: Jan-Ole Hesselberg, janohes@student.sv.uio.no, janoleh@gmail.com.

Editorial group: Cochrane Methodology Review Group.
Publication status and date: New, published in Issue 11, 2020.

Citation: Hesselberg J-O, Dalsbø TK, Stromme H, Svege I, Fretheim A. Reviewer training for improving grant and journal peer review (Protocol). *Cochrane Database of Systematic Reviews* 2020, Issue 11. Art. No.: MR000056. DOI: 10.1002/14651858.MR000056.

ABSTRACT

Objectives

This is a protocol for a Cochrane Review (methodology). The objectives are as follows:

To evaluate the effect of peer reviewer training on the quality of grant and journal peer review.

BACKGROUND

Both research funders and scientific journals use peer review for deciding which projects to fund or articles to publish. Most funders and journals are likely to have a set of criteria they want the project proposals or articles to be reviewed by, and most are likely to have formulated instructions for how to do this. To increase compliance with the defined review criteria some funders and journals train their reviewers. This is done through a variety of interventions and methods, including (but not limited to) training sessions, courses, handbooks, written instructions, feedback and guidance. However the training is delivered, the primary goal of these training interventions is to improve the quality of the peer review.

Different outcome measures are used to evaluate the quality of peer review. Some include how well the research project or article performs in terms of different bibliometrics, stakeholders' evaluation of peer reviews, inter-reviewer agreement and authors' adherence to guidelines.

Several studies show that both grant and journal peer review performs suboptimally on several important outcomes measures (Bornmann 2011; Guthrie 2017). Reviewer training is used as an intervention to improve these outcomes. However, studies on the effects of such training yield inconsistent results (Bruce 2016), and there is no up-to-date systematic review addressing this question (see Why it is important to do this review).

Description of the methods being investigated

According to Elsevier, one of the world's largest scholarly publishers, "Reviewers evaluate article submissions to journals based on the requirements of that journal, predefined criteria, and the quality, completeness and accuracy of the research presented". (Elsevier 2018).

Both funders and journals are likely to have a set of formal documents describing the aims and the scope of their work (e.g. a mission statement), the specific criteria the funding applications and articles will be measured against (e.g. guidelines, call for proposals) and specifics on the scales used for scoring and how the criteria should be scored (e.g. peer reviewer scoring instructions). Peer reviewer training is likely to address the aims of the funder or journal, the peer review criteria and scoring instructions, with the goal of increasing the quality of the peer review through securing adherence to them.

Surveys of peer reviewers show that they get little formal training in peer review, that they would like more guidance and that they think it will improve the quality of the peer review (Mulligan 2013; Sense About Science 2009; Warne 2016). When training is done, it is delivered in a variety of ways. Some use passive training strategies, such as guidelines, written instructions, handbooks or videos. Others use more active strategies, such as live training sessions, online courses or mentoring. These strategies are likely to be implemented in advance of the peer review but some also use post-review training strategies, such as feedback or evaluations to train peer reviewers for their next peer review session.

How these methods might work

The training of peer reviewers might improve the peer review process through several mechanisms.

The peer reviewers' interpretation of criteria and the weighting of them have been shown to vary (Abdoul 2012), and the tasks viewed as critical to peer reviewers are often not congruent with the tasks requested by the funding program officers and journal editors (Chauvin 2015). These discrepancies might lower the inter-rater and intra-rater reliability, and training might improve this.

Human judgment and decision-making processes are known to be biased, and research suggests this might also be the case in grant and journal peer review (Bornmann 2006; Gallo 2018; Langfeldt 2006; Lee 2013; Walker 2015). Factors such as cognitive distance (Wang 2015), halo effects, leniency, anchor effects and sequential contrast effects have been shown to influence several different assessment settings (Bhargava 2014; Danziger 2011; Olbrecht 2010; Sattler 2015), and informing peer reviewers about the risk of biases and training in how to avoid them might reduce their effects.

Furthermore, in some instances peer review is done in groups and social psychological phenomenons such as groupthink, group polarization or bandwagon effects might influence decisions in a biased way, particularly in panel discussions and applicant interviews (Olbrecht 2010). Training in how to avoid these pitfalls might increase the reliability and validity of peer review.

Why it is important to do this review

In late 2014, there were about 34,550 scholarly peer-reviewed journals in the world, publishing approximately 2.5 million articles that year (Ware 2015). The editorial decision to publish or reject manuscripts submitted to these journals relies heavily on peer review and has significant consequences for both researchers and research output. Researchers are likely to be evaluated on both how often and in which journals they publish when they apply for jobs or grants, while research output is affected not only by the peer reviewers' recommendations regarding publication, but also their suggestions regarding revisions to the manuscript, including the methodology of the study.

Worldwide, large funds are distributed through grant application processes in which experts and peers consider whether or not a project is worthy of support. According to a recent review, "peer review decisions award >95% of academic medical research funding" (Guthrie 2017). In the USA, the National Institutes of Health alone invests more than $40 billion a year in medical research (National Institutes of Health 2020), and more than 80% of this is awarded through competitive grants. In the EU, nearly €80 billion will be granted through the 'Horizon 2020 programme', and in the first three years of that initiative more than 20,000 peer reviewers were involved (European Commision 2018). As with journal peer review, the consequences are significant. Not only does the peer review affect the distribution of research funds, success in grant application also affects researchers' future success chances (Bol 2018).

Still, little seems to be known about the effect of peer reviewer training on the quality of peer review. No systematic reviews have studied the effects of this training in grant peer review (Guthrie 2017; Sattler 2015), and the most recent review of studies on

reviewer training in journal peer review searched for studies in June 2015 (Bruce 2016).

Furthermore, grant and journal peer review are very similar processes, often using overlapping criteria (such as methodological quality, impact and originality) and including both of these types of peer review should add power to this systematic review. Two Cochrane Methodology reviews are related to journals (Jefferson 2007) and grant peer review (Demicheli 2007), respectively. They focus on peer review as a measure for improving the quality of funded research and study reports, and they both highlight important challenges with peer review. As our review focuses on training as a measure to address some of these challenges, we believe it will complement the two existing reviews.

OBJECTIVES

To evaluate the effect of peer reviewer training on the quality of grant and journal peer review.

METHODS

Criteria for considering studies for this review

Types of studies

We will include randomized trials and cluster-randomized trials comparing training interventions with usual processes, no training interventions or other interventions to improve the quality of peer review.

Types of data

Peer reviewers in research grants and the scientific publication process.

Types of methods

Any intervention aimed at training peer reviewers. Based on surveys of peer reviewers and research on the effects of training, the most common training interventions are likely to include:

- guidelines, instructions, checklists and templates;
- guidance or mentoring;
- feedback and evaluation;
- workshops, seminars and webinars;
- self-administered online/video courses.

The interventions might be journal- or funder-specific, or aimed at peer review more generally. Workshop, seminars and webinars are usually held in groups. Some of the interventions are more time consuming than others. Feedback and evaluation are typically given after, or during the peer review process.

Types of outcome measures

Our main outcome of interest is the quality of the review, however measured. Based mainly on a recent systematic review of interventions to improve journal peer review (Bruce 2016), we expect the outcome measures below. We expect that both objective and subjective measures will be reported and we will analyze both.

Primary outcomes

The following outcomes are considered to be primary measures of peer review quality.

- Completeness of reporting in articles based on relevant guidelines, such as CONSORT (Altman 2001), TIDieR (Hoffmann 2014) and SPIRIT (Chan 2013)
- Peer reviewer identification of deliberate, inserted flaws in manuscripts and proposals

Secondary outcomes

The following secondary outcomes are considered relevant.

- Bibliometric scores (such as citation rates, Altmetrics score and save rate)
- Evaluations of peer reviewers or their reviews by stakeholders, such as grant administration or editors (e.g. scored through use of instruments like the Review Quality Instrument (van Rooyen 1999) and The Manuscript Quality Assessment Instrument (Goodman 1994))
- Degree of agreement between peer reviewers (changes in inter-rater reliability measures like weighted kappa and intra-class correlation)
- Process-centred outcomes (such as speed or cost of reviewing)
- Peer reviewer satisfaction with the review process
- Completion rate and speed of funded projects (applicable only to grant peer review)

Studies will be included based on the eligibility criteria above, regardless of whether or not they include the outcome measures mentioned.

Search methods for identification of studies

We will search for all published and unpublished studies of the effect of peer reviewer training on the quality of peer review, without restrictions on language or publication status other than those arising because of the sources we will search. The search strategy will be developed by an information specialist (HS) and peer reviewed by another information specialist.

Electronic searches

We will search MEDLINE with the strategy presented in Appendix 1. The strategy will be adapted for the following databases.

- Cochrane Central Register of Controlled Trials (CENTRAL)
- Embase
- PsycINFO
- CINAHL (Cumulative Index to Nursing and Allied Health)
- ERIC
- ProQuest Dissertations & Theses Global™
- Web of Science
- OpenGrey

We will also search the following trials registries for planned and ongoing trials:

- ClinicalTrials.gov
- WHO International Clinical Trials Registry Platform (ICTRP)
- Open Science Framework (OSF)

Searching other resources

We will check the reference list of included studies and any relevant systematic reviews for references to relevant trials (Horsley 2011).

We will contact major research funding agencies and researchers who are known or expected to have conducted relevant research and ask them for information on relevant trials.

Data collection and analysis

Selection of studies

An information specialist (HS) will conduct the searches and remove duplicates, usingEndNote X9.

Two review authors (JOH, ICS) will assess the studies independently in three steps. First, the studies will be screened by title and abstracts. Second, the records identified as potentially eligible in step one will be screened by reading full-text articles and study protocols. Third, the results from the two independent screenings will be compared. In case of disagreements, the studies will be assessed by a third author (AF). We will make the final decision of whether to include or exclude a study in a face-to-face or online meeting between the three authors. We will use Covidence (Covidence) for the screening process.

We will complete a PRISMA flow diagram (Moher 2015) and 'Characteristics of excluded studies' table and include these in the final publication.

Data extraction and management

One review author (JOH) will extract general information from the studies (see bullet points below). Two review authors (JOH and TD) will extract the remaining data independently using a version of the Cochrane data extraction template (all studies), which we will pilot on at least two studies. In case of disagreements, studies will be assessed by a third review author (AF) blinded to the details of the disagreement. We will make the final decision in a face-to-face or online meeting between the three review authors.

We will extract the following study characteristics.

- General information: author details, year and language of publication
- Methods: study design, study setting, withdrawals, total duration of the study, date of study
- Participants: numbers and proportions in groups, peer reviewer experience, field of expertise, gender, language, country of residence
- Interventions: detailed description of the intervention, comparison, how the intervention was designed and by whom, delivery format, temporal length of intervention, who delivers the intervention
- Outcomes: specified primary and secondary outcomes and relevant outcomes identified, reported time points
- Notes: study funding, conflicts of interest

We will use the TIDieR checklist to describe the components of the intervention (Hoffmann 2014), and we will collect characteristics of the included studies in sufficient detail to populate a table of 'Characteristics of included studies'.

Assessment of risk of bias in included studies

Two review authors (JOH and TD) will assess the risk of bias independently for each study included in the review. We will use the Cochrane 'Risk of bias' tool. The assessments will be made in Covidence and transferred to Review Manager (RevMan 2020). If information is not available in the published reports and clarification is needed, we will contact study authors or funders. In case of disagreements in assessing the risk of bias, a third review author (AF) will assess this. We will make the final decision in a face-to-face or online meeting between the three review authors.

We will assess the risk of bias across the following domains.

- Random sequence generation and allocation concealment (allocation bias)
- Blinding of personnel, participants (performance bias) and outcome assessors (detection bias)
- Incomplete outcome data (attrition bias)
- Selective reporting

Following the *Cochrane Handbook of Systematic Reviews of Interventions* (Higgins 2011), we will judge each study to be at high, unclear or low risk of bias. We will present results of these assessments in a Rrisk of bias' table and 'Risk of bias' summary' figure.

Measures of the effect of the methods

We will collect all outcomes reported in the studies along with how they were measured (self-report, chart-abstraction, other objective primary or secondary outcome). For included outcomes, we will extract the intervention effect estimates reported by the investigators of the study, along with its confidence interval and the method of statistical analysis used to calculate it.

If more than one adherence outcome is reported within the same study, we will use the primary outcome as defined by the study authors. If a primary outcome is not clearly defined, we will calculate and use the median value from all relevant outcomes. We will request additional information from the authors if reports do not contain sufficient data (Young 2011).

Continuous outcomes

For continuous outcomes, we will analyze data based on the mean, standard deviation (SD) and the number of people assessed for both the intervention and comparison groups to calculate the mean difference (MD) and 95% confidence interval (CI). In studies where different measures have been used to assess the impact of the intervention on the same outcome, we will use the standardized mean difference (SMD) with its 95% CI.

Dichotomous outcomes

For dichotomous outcomes, we will analyze data based on the number of events and the number of people assessed in the intervention and comparison groups. We will use these to calculate the risk ratio (RR) and 95% CI.

Cluster-randomized trials

If cluster-randomized trials are identified the interventions are likely to be allocated to groups of peer reviewers. We will analyze such trials using the average cluster size and the value of the intraclass correlation coefficient (ICC). We will follow the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).
If ICC estimates are not reported, we will contact trial authors for these, but if they are not available, we will use an estimate from a

similar trial. Sensitivity analyses will be conducted if trials have not accounted for clustering.

Studies with multiple intervention groups

If multiple intervention groups are reported, we will include only the relevant groups. We will exclude or combine arms if we are conducting meta-analyses with pairwise comparisons, and we will do this based on our judgement.

Dealing with missing data

We will asses each study for missing data and attrition. If data are missing or unclear, we will contact the study authors and funders, to request these missing data or the reasons for and characteristics of dropouts (Young 2011).

Assessment of heterogeneity

We will assess statistical heterogeneity by inspecting the forest plot visually for overlap of confidence intervals and for outliers. Lack of overlap and outliers will be interpreted as possible heterogeneity. We will use the Chi^2 test to assess statistical heterogeneity (Deeks 2017), with the significance level set at $P < 0.10$. We will also use the I^2 statistic (Higgins 2003) and will interpret results above 30% as possible heterogeneity. We will also use subgroup and sensitivity analyses to assess heterogeneity.

Assessment of reporting biases

If our review includes 10 or more included studies, we will investigate publication (or other small-study) bias by visually inspecting funnel plots for skewness. Skewness/asymmetry will be further investigated by using Egger's test for continuous outcomes (Egger 1997) and the Harbord's test for dichotomous outcomes (Harbord 2006). If asymmetries are detected, we will discuss possible explanations and consider performing sensitivity analyses.

Data synthesis

If the data quality is adequate and the included trials are sufficiently similar in terms of interventions, comparisons, participants, settings and outcome measures, we will meta-analyze the data to provide an overall effect estimate. We are not aware of evidence suggesting that the effect of the different training interventions will have markedly different effects. We will therefore presume similar effects for all training interventions, across both grant and journal peer review, and will meta-analyze the data collectively. We will use Review Manager 5.4 for data analysis (RevMan 2020).

Based on the expected variability in samples and interventions, we will use a random-effects model to incorporate this heterogeneity. For continuous variables, we will use the inverse-variance method. For dichotomous variables, we will use the method proposed by Mantel-Haenszel (Mantel 1959).
If several studies measured the same outcomes but using different tools, we will calculate the standardized mean difference (SMD) and 95% CI using the inverse variance method in RevMan.

'Summary of findings' table

We will create a 'Summary of findings' table showing the following five outcomes.

- Completeness of reporting in articles based on guidelines
- Peer reviewer identification rate of deliberate, inserted flaws in manuscripts and proposals
- Bibliometric scores
- Evaluations of peer reviewers or peer reviews by stakeholders, such as grant administration or editors
- Degree of agreement between peer reviewers

We will use the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness, and publication bias) to assess the certainty of the evidence as it relates to the studies which contribute data to the meta-analyses for the prespecified outcomes (Guyatt 2008). We will use methods and recommendations described in section 8.5 and chapter 12 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2017). We will justify all decisions to downgrade the evidence using footnotes, and we will make comments to aid the reader's understanding of the review, where necessary.

Subgroup analysis and investigation of heterogeneity

If heterogeneity is above the thresholds described in Assessment of heterogeneity, we may conduct subgroup analyses to investigate possible sources of heterogeneity. The following subgroups will be analyzed if appropriate data are available.

Type of peer review

Even though several peer review criteria overlap between grants and journals (e.g. methodological quality, originality and impact), there are also important differences. For example, in grant peer review, the expected feasibility of the study and the merits of the research environment for the proposed research are essential criteria that are not a part of journal peer review. This might result in increased heterogeneity.

Degree of involvement

Training interventions that demand more involvement from the peer reviewers are expected to yield better outcomes than those that demand little. Ideally, time spent in training should be the basis of a subgroup analysis. However, we are doubtful that a sufficient number of studies will have precise measures of this. As an alternative, we will analyze types of training that are likely to demand different degrees of involvement from the peer reviewers.

Training expected to demand a higher degree of involvement are:

- personal mentoring/guidance/feedback/evaluation;
- seminars/workshops/webinars.

Training expected to demand a lower degree of involvement are:

- guidelines and written or verbal instructions;
- checklists/template.

Sensitivity analysis

We will conduct a sensitivity analysis if:

- cluster-randomized trials have not adjusted for clustering (i.e. sensitivity analysis where we exclude data from non-adjusted cluster-randomized trials);
- significant heterogeneity is detected (forest plots will be inspected to determine possible sources);

- allocation bias is detected (studies deemed at high risk of bias will be excluded).

ACKNOWLEDGEMENTS

We wish to thank information specialist Marit Johansen at the Norwegian Institute of Public Health for reviewing the search strategy and offering important suggestions. We also wish to thank secretary general Hans Christian Lillehagen at the Norwegian Foundation Dam for providing feedback on and support to the project in its initial stages. His feedback was of particular importance to the selection of criteria for considering studies for the review.

REFERENCES

Additional references

Abdoul 2012

Abdoul H, Perrey C, Amiel P, Tubach F, Gottot S, Durand-Zaleski I, et al. Peer review of grant applications: criteria used and qualitative study of reviewer practices. *PLOS One* 2012;**7**(9):e46054.

Altman 2001

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine* 2001;**134**(8):663-94.

Bhargava 2014

Bhargava S, Fisman R. Contrast effects in sequential decisions: evidence from speed dating. *Review of Economics and Statistics* 2014;**96**(3):444-57.

Bol 2018

Bol T, de Vaan M, van de Rijt A. The Matthew effect in science funding. *Proceedings of the National Academy of Sciences of the United States of America* 2018;**115**(19):4887-90.

Bornmann 2006

Bornmann L, Daniel H-D. Potential sources of bias in research fellowship assessments: effects of university prestige and field of study. *Research Evaluation* 2006;**15**(3):209-19.

Bornmann 2011

Bornmann L. Scientific peer review. *Annual Review of Information Science and Technology* 2011;**45**(1):197-245.

Bruce 2016

Bruce R, Chauvin A, Trinquart L, Ravaud P, Boutron I. Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. *BMC Medicine* 2016;**14**(1):85.

Chan 2013

Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Annals of Internal Medicine* 2013;**158**(3):200-7.

Chauvin 2015

Chauvin A, Ravaud P, Baron G, Barnes C, Boutron I. The most important tasks for peer reviewers evaluating a randomized controlled trial are not congruent with the tasks most often requested by journal editors. *BMC Medicine* 2015;**13**:158.

Covidence [Computer program]

Veritas Health Innovation Covidence. Version accessed 29 May 2019. Melbourne, Australia: Veritas Health Innovation. Available at covidence.org.

Danziger 2011

Danziger S, Levav J, Avnaim-Pesso L. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America* 2011;**108**(17):6889-92.

Deeks 2017

Deeks JJ, Higgins JP, Altman DG (editors) on behalf of the Cochrane Statistical Methods Group. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JP, Churchill R, Chandler J, Cumpston MS (editors), Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017). Cochrane, 2017. Available from www.training.cochrane.org/handbook.

Demicheli 2007

Demicheli V, Di Pietrantonj C. Peer review for improving the quality of grant applications. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art. No: MR000003. [DOI: 10.1002/14651858.MR000003.pub2]

Egger 1997

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)* 1997;**315**(7109):629-34.

Elsevier 2018

Elsevier. Role of a reviewer. https://www.elsevier.com/reviewers/role (accessed 30 July 2020).

European Commision 2018

European Commision. HORIZON 2020 in full swing. Three years on. Key facts and figures 2014-2016. retrieved from https://ec.europa.eu/research/mariecurieactions/sites/mariecurie2/files/h2020_threeyearson_a4_horizontal_2018_web.pdf 2018.

Gallo 2018

Gallo S, Thompson L, Schmaling K, Glisson S. Risk evaluation in peer review of grant applications. *Environment Systems and Decisions* 2018;**38**(2):216-29.

Goodman 1994

Goodman SN. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Annals of Internal Medicine* 1994;**121**(1):11-21. [DOI: 10.7326/0003-4819-121-1-199407010-00003]

Guthrie 2017

Guthrie S, Ghiga I, Wooding S. What do we know about grant peer review in the health sciences? *F1000Research* 2017;**6**:1335.

Guyatt 2008

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical Research Ed.)* 2008;**336**(7650):924-6.

Harbord 2006

Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 2006;**25**(20):3443-57.

Higgins 2003

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.)* 2003;**327**(7414):557-60.

Higgins 2011

Higgins JP, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2017

Higgins JP, Altman DG, Sterne JA. Chapter 8: Assessing risk of bias in included studies. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated June 2017). The Cochrane Collaboration, 2017. Available from www.training.cochrane.org/handbook.

Hoffmann 2014

Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ (Clinical Research Ed.)* 2014;**348**:g1687.

Horsley 2011

Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. *Cochrane Database of Systematic Reviews* 2011, Issue 8. Art. No: MR000026. [DOI: 10.1002/14651858.MR000026.pub2]

Jefferson 2007

Jefferson T, Rudin M, Brodney Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art. No: MR000016. [DOI: 10.1002/14651858.MR000016.pub3]

Langfeldt 2006

Langfeldt L. The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation* 2006;**15**(1):31-41.

Lee 2013

Lee CJ, Sugimoto CR, Zhang G, Cronin B. Bias in peer review. *Journal of the American Society for Information Science and Technology* 2013;**64**(1):2-17.

Mantel 1959

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *JNCI: Journal of the National Cancer Institute* 1959;**22**(4):719–48.

Moher 2015

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 2015;**4**(1):1. [DOI: 10.1186/2046-4053-4-1]

Mulligan 2013

Mulligan A, Hall L, Raphael E. Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology* 2013;**64**(1):132-61.

National Institutes of Health 2020

National Institutes of Health. Budget. https://www.nih.gov/about-nih/what-we-do/budget (accessed 30 July 2020).

Olbrecht 2010

Olbrecht M, Bornmann L. Panel peer review of grant applications: what do we know from research in social psychology on judgment and decision-making in groups? *Research Evaluation* 2010;**19**(4):293-304.

RevMan 2020 [Computer program]

The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.4. Copenhagen: The Cochrane Collaboration, 2020.

Sattler 2015

Sattler DN, McKnight PE, Naney L, Mathis R. Grant peer review: improving inter-rater reliability with training. *PLOS One* 2015;**10**(6):e0130450.

Sense About Science 2009

Sense About Science. Peer Review Survey 2009: Full Report. http://senseaboutscience.org/activities/peer-review-survey-2009/ (accessed 30 July 2020).

van Rooyen 1999

van Rooyen S, Black N, Godlee F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. *Journal of Clinical Epidemiology* 1999;**52**(7):625-9.

Walker 2015

Walker R, Barros B, Conejo R, Neumann K, Telefont M. Personal attributes of authors and reviewers, social bias and the outcomes of peer review: a case study. *F1000Research* 2015;**4**:21.

Wang 2015

Wang Q, Sandström U. Defining the role of cognitive distance in the peer review process with an explorative study of a grant scheme in infection biology. *Research Evaluation* 2015;**24**(3):271-81.

Ware 2015

Ware M, Mabe M. The STM Report - an overview of scientific and scholarly journal publishing. Retrieved from https://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf 2015.

Warne 2016

Warne V. Rewarding reviewers - sense or sensibility? A Wiley study explained. *Learned Publishing* 2016;**29**(1):41-50.

Young 2011

Young T, Hopewell S. Methods for obtaining unpublished data. *Cochrane Database of Systematic Reviews* 2011, Issue 11. Art. No: MR000027. [DOI: 10.1002/14651858.MR000027.pub2]

APPENDICES

Appendix 1. MEDLINE search strategy

1. Peer Review/
2. Peer Review, Research/
3. peer review*.tw,kf.
4. or/1-3
5. Education/
6. Education, Professional/
7. exp Inservice Training/
8. Teaching/
9. Mentors/
10. Mentoring/
11. train*.tw,kf.
12. workshop*.tw,kf.
13. school*.tw,kf.
14. feedback.tw,kf.
15. mentor*.tw,kf.
16. coach*.tw,kf.
17. teach*.tw,kf.
18. taught.tw,kf.
19. educat*.tw,kf.
20. exercis*.tw,kf.
21. guid*.tw,kf.
22. instruct*.tw,kf.
23. practice.tw,kf.
24. handbook*.tw,kf.
25. manual*.tw,kf.
26. course*.tw,kf.
27. procedure*.tw,kf.
28. program*.tw,kf.
29. assessment*.tw,kf.
30. evaluation*.tw,kf.
31. checklist*.tw,kf.
32. check-list*.tw,kf.
33. correspond*.tw,kf.
34. template*.tw,kf.
35. or/5-34
36. 4 and 35
37. randomized controlled trial.pt.
38. controlled clinical trial.pt.
39. (randomized or randomised).ti,ab.
40. placebo.ab.
41. clinical trials as topic.sh.
42. randomly.ab.
43. trial.ti.
44. or/37-43
45. exp animals/ not humans.sh.
46. 44 not 45
47. 36 and 46

HISTORY

Protocol first published: Issue 11, 2020

CONTRIBUTIONS OF AUTHORS

JOH, HS, TD, IS, and AF developed the protocol.
HS will search for the studies and obtain copies of them.
JOH and IS will select studies to include, with AF serving as the arbitrator.
JOH and TD will extract data from the studies.
JOH will enter data into RevMan.
JOH and TD will carry out and interpret the analysis.
JOH will draft the final review.
All authors will contribute to the final version of the review

DECLARATIONS OF INTEREST

JOH: none known. JOH is the chief program officer at The Norwegian ExtraFoundation (ExtraStiftelsen), a research funder that also funded the PhD which this systematic review is a part of.
IS: none known. IS is an advisor at The Norwegian ExtraFoundation (ExtraStiftelsen).
TD: none known.
HS: none known.
AF: none known.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- Stiftelsen Dam, Norway

  This systematic review is a part of a PhD that is funded by The Norwegian Foundation Dam (Stiftelsen Dam). Jan-Ole Hesselberg is the chief program officer and Ida C. Svege is a senior advisor at the foundation.