Université Paris Descartes

**Ecole doctorale 393 Pierre Louis de Santé Publique : Epidémiologie et Sciences de l'Information Biomédicale**

*Centre de recherche en épidémiologie et statistiques Sorbonne Paris Cité*
*Equipe Méthodes en évaluation thérapeutique des maladies chroniques (METHODS)*

# Développement et évaluation d'interventions visant à améliorer le processus de « peer-review » en recherche biomédicale

Par Anthony Chauvin

**Thèse de doctorat d'Epidémiologie**

Dirigée par Isabelle BOUTRON et co-dirigée par David MOHER

**Présentée et soutenue publiquement le 20 septembre 2018**

**Devant un jury composé de** :

Monsieur le Professeur Bruno GIRAUDEAU, PU-PH, Rapporteur
Monsieur le Professeur Yonathan FREUND, PU-PH, Rapporteur
Monsieur le Professeur RICARD Jean-Damien, PU-PH, Examinateur
Monsieur le Professeur ADNET Frédéric, PU-PH, Examinateur
Madame le Professeur Isabelle BOUTRON, PU-PH, Directrice de thèse

# Remerciements

Avant tout je tiens à remercier **Madame le Professeur Isabelle Boutron** qui a, dès le Master 2 accepté de me guider dans ces démarches de recherche. Je suis honoré qu'elle ait accepté de diriger également ma Thèse d'Université. Sa rigueur, sa passion et son expérience sont un exemple pour moi. J'espère avoir été à la hauteur de sa confiance et pouvoir continuer à collaborer avec elle durant de nombreuses années.

Merci au **Professeur Philippe Ravaud** pour son accueil au sein du centre du CRESS.

Un merci tout particulier au **Professeur David Moher** qui m'a accompagné tout au long de mon travail de Thèse.

Je remercie aussi le **Professeur Bruno Giraudeau** et le **Professeur Yonathan Freund** d'avoir accepté d'être rapporteurs de ce travail

Je remercie le **Professeur Jean-Damien Ricard** et le **Professeur Frederic Adnet** de m'avoir fait l'honneur d'être membre du jury.

Merci à tous les membres de l'équipe METHODS et du CRESS : Audrey, Isabelle, Aurélie, Agnès, Amélie, Perrine, Romana, Aymar, Ignacio, Michael, Céline, Aïda, Alexandre.

Merci à mon Maître et Chef de Service, **Monsieur le Professeur Patrick Plaisance**, qui a su placer sa confiance en moi.

Une pensée toute particulière à mes partenaires de bureau qui ont su me supporter durant toutes ces années : Stéphanie Sidorkiewicz et bien sûr Youri Yordanov. J'espère que la passion de la recherche continuera à nous animer durant de nombreuses années et que nous aurons toujours autant de discussions passionnées !!

Merci à ma femme Céleste, mon roc qui chaque jour doit me supporter… Sans toi je n'en serais jamais arrivé là. Merci pour ta clairvoyance. A mes trois enfants, mes bulles d'oxygène et mes sources de motivation.

A mes parents et ma soeur, merci de toujours être présents à nos côtés et d'avoir une oreille attentive.

A mon frère….

Aux amis « Lyonnais » toujours présents : Fredy, Emilie et Thomas. Et bien sûr une pensée à toute la meute de Boubou.

## Résumé

Le processus de « peer-review » est considéré comme la pierre angulaire de la publication scientifique. Or, des travaux ont démontré que les évaluations des peer-reviewers sont souvent discordantes, qu'ils ne parviennent pas à détecter la fraude, les erreurs et que leur évaluation est influencée par les résultats.

L'objectif de ce travail de thèse était d'explorer et de repenser le processus de la peer-review éditoriale.

Dans un premier temps, nous avons identifié les tâches attendues des peer-reviewers lors de l'évaluation d'un manuscrit rapportant un essai contrôlé randomisé et, comparé l'importance de ces tâches selon le point de vue des peer-reviewers et des éditeurs. Notre étude a permis de démontrer que les tâches les plus importantes pour les peer-reviewers n'étaient pas conformes aux tâches les plus souvent demandées par les éditeurs.

Dans un deuxième temps, nous avons réalisé une revue systématique ainsi qu'une méta-analyse des essais contrôlés randomisés évaluant une intervention visant à améliorer la qualité de la peer-review dans les journaux biomédicaux. Nous avons identifié 22 essais contrôlés randomisés qui ont évalué 5 interventions. Ces essais incluaient un faible nombre de participants, étaient pour la majorité monocentriques, et avaient des biais qui limitaient l'interprétation de leurs résultats.

Enfin, nous avons développé et évalué un outil d'aide à la réalisation de la peer-review d'essais randomisés, associé à une formation en ligne. Cet outil vise spécifiquement à évaluer la qualité de la présentation des essais, c'est-à-dire le respect des recommandations du CONSORT statement (i.e., recommandations internationales exigées par les éditeurs pour rapporter les essais randomisés de façon complète et transparente). Nous avons comparé la performance de jeunes peer-reviewers formés utilisant l'outil, avec la performance de reviewers ayant évalué le manuscrit dans le cadre du processus de peer-review classique. Nous avons inclus 120 essais contrôlés randomisés. Nos résultats ont mis en évidence que l'outil a amélioré le nombre d'items correctement classés par manuscrit par les jeunes peer-reviewers. Cette constatation suggère que l'utilisation d'une checklist de type CONSORT se focalisant sur certains items précis pourrait améliorer la qualité du processus de peer-review, et donc la qualité du manuscrit publié.

A l'issue de ce travail, nous proposons de repenser le processus de peer-review qui pourrait être réalisé en 2 étapes avec : des peer-reviewers juniors évaluant le respect des recommandations du CONSORT statement et des reviewers expérimentés évaluant la méthodologie et la pertinence de la question.



**Mots-Clés** : peer-review ; essais contrôlés randomisés ; outil informatique ; CONSORT Statement

## Abstract

The peer-review process is a cornerstone of biomedical research publication. However, studies have shown that peer-reviewers' evaluations are often discordant, that they fail to detect fraud, errors and that their evaluation is influenced by the results.

The objective of this thesis work was to explore and rethink the process of peer-review editorial.

As a first step, we identified the tasks expected of peer-reviewers when evaluating a randomized controlled trial report and compared the importance of these tasks from the points of view of peer-reviewers and editors. Our study showed that the most important tasks for peer reviewers were not congruent with the tasks most often requested by journal editors in theirs guidelines to reviewers.

In a second step, we conducted a systematic review and a meta-analysis of randomized controlled trials evaluating an intervention aimed at improving the quality of peer-review in biomedical journals. We identified 22 randomized controlled trials that evaluated 5 interventions. These trials included a small number of participants, were mostly monocentric and had biases that limited the interpretation of their results.

Finally, we have developed and evaluated a tool to help achieve the peer-review of randomized trials associated with online training. This tool is specifically aimed at evaluating the quality of the reporting that means the compliance with the recommendations of the CONSORT statement (i.e., international recommendations required by publishers to report randomized trials in a complete and transparent manner). We compared the performance of young trained peer-reviewers using the tool with the performance of reviewers evaluating the manuscript as part of the usual peer-review process. We included 120 randomized controlled trials.

Our results showed that the tool improved the number of items correctly classified per manuscript by young peer-reviewers. This finding suggests that the use of a CONSORT checklist focusing on specific items could improve the quality of the peer-review process, and thus the quality of the published manuscript

At the end of this work, we propose to rethink the peer-review process that could be carried out in 2 stages with junior peer-reviewers assessing compliance with the recommendations of the CONSORT statement and experienced reviewers evaluating the methodology and relevance of the question.



**Keywords:** peer-review; randomized controlled trials; computer tool; CONSORT Statement

## Laboratoire d'accueil

Centre de Recherche Epidémiologie et Statistiques Sorbonne Paris Cité (CRESS)

CRESS U1153, équipe METHODS : Méthodes en Évaluation Thérapeutique des Maladies Chroniques

Centre d'Épidémiologie Clinique

Hôpital Hôtel Dieu

1 place du parvis Notre Dame

75004 Paris, France

## Publications liées à la thèse

Ce travail de doctorat s'appuie sur les articles suivants :

**Publications en rapport avec la thèse**

- **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 Med. 2015 Jul 3;13:158
- 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 Jun 10;14(1):85
- **Chauvin A**, Moher D, Altman D, Schriger DL, Alam S, Hopewell S, Shanahan DR, Recchioni A, Ravaud P, Boutron I. A protocol of a cross-sectional study evaluating an online tool for early career peer reviewers assessing reports of randomised controlled trials. BMJ Open. 2017 Sep 15;7(9):e017462
- **Chauvin A**, Moher D, Altman D, Schriger DL, Alam S, Hopewell S, Shanahan DR, Recchioni A, Ravaud P, Boutron I. A cross-sectional study evaluating an online tool for early career peer reviewers assessing reports of randomised controlled trials. En cours de redaction

∗ Co- premier auteur

# Liste des principales abréviations

IC : intervalle de confiance

SMD : différence moyenne standardisée

CONSORT : CONsolidated Standards Of Reporting Trials

# Table des matières

1. Introduction ..... 11
1.1 La peer review ou relecture par les pairs ..... 11
1.1.1 Définition ..... 11
1.1.2 Historique ..... 11
1.1.3 Les objectifs de la peer review ..... 14
1.1.4 Le processus de peer review ..... 14
1.1.5 Les différentes modalités de la peer review ..... 16
1.1.5.1 L'évaluation pré publication ..... 16
1.1.5.2 L'évaluation post publication ..... 17
1.1.6 Les critiques de la peer review ..... 18
1.1.6.1 Coût et délai ..... 18
1.1.6.2 Reproductibilité ..... 19
1.1.6.3 Présentation inadapté des manuscrits ..... 20
1.1.6.4 Détection des spins. ..... 21
1.1.6.5 Changement du critère de jugements ..... 22
1.1.6.6 Détection des erreurs ..... 22
1.1.6.7 Détection des conflits d'intérêts ..... 23
1.1.6.8 Les biais cognitifs ..... 25
1.1.6.8.1 Biais d'objectivité ..... 25
1.1.6.8.2 Biais de genre ..... 25
1.1.6.9 Conduites inappropriées ..... 26
1.1.6.9.1 Plagiat ..... 26
1.1.6.9.2 Fausse peer-review ..... 27
2. Objectifs de la thèse ..... 28
3. Les tâches incombant aux peer-reviewers lors de l'évaluation d'un essai contrôlé randomisé 29
3.1 Justification ..... 29
3.2 Objectifs ..... 29
3.3 Méthode ..... 30
3.4 Résultats ..... 31
3.5 Discussion et limite ..... 34
3.6 Conclusion ..... 35
3.7 Article ..... 36
3.8 Appendices de l'article ..... 47
4. Les interventions visant à améliorer la qualité de l'évaluation par les peer-reviewers des revues biomédicales ..... 72
4.1 Justification ..... 72
4.2 Objectifs ..... 72
4.3 Méthode ..... 72
4.4 Résultats ..... 73
4.5 Discussion et limite ..... 74
4.6 Conclusion ..... 75
4.7 Article ..... 76
4.8 Appendices de l'article ..... 78
5. L'outil COBPeer ..... 133
5.1 Justification de l'étude ..... 133
5.2 Le CONSORT: *Consolidated Standards of Reporting Trials* ..... 134

5.3 Objectif ..... 136
5.4 Méthodes ..... 137
5.4.1 Développement de l'outil COBPeer ..... 137
5.4.2 Intégralité du rapport des items pour les essais contrôlés randomisés ..... 139
5.5 Intervention ..... 139
5.6 Critères de jugements principaux ..... 143
5.7 Résultats ..... 143
5.8 Discussion et limites ..... 146
5.9 Article ..... 147
5.10 Appendices de l'article ..... 158
6. Discussion et perspectives ..... 160
6.1 Principaux résultats ..... 160
6.2 Discussion ..... 161
6.3 Perspectives ..... 163
6.3.1 Une peer-review en deux étapes ..... 163
6.3.2 Autres formes de peer-review ..... 163
7. Conclusion ..... 165

Bibliographie ..... 166

# Liste des figures

Figure 1. Les différentes étapes du processus de la peer-review ..... 16

Figure 2. Niveau des spins dans la conclusion des résumés, avant et après peer-review ..... 21

Figure 3. Proportion de participants évaluant la tâche dans le premier tertile (c-à-d≥2) et la proportion d'éditeurs explicitant la tâche dans leurs recommandations aux auteurs ..... 33

Figure 4. Exemple d'exercice proposé aux participants ..... 138

Figure 5. Design de l'étude COBPeer ..... 141

Figure 6. Aperçu de l'outil COBPeer pour l'évaluation des essais contrôlés randomisés ..... 142

# 1. Introduction

## 1.1 La peer-review éditoriale ou relecture par les pairs

### 1.1.1 Définition

Le processus de révision par les pairs ou peer-review est un travail collaboratif d'experts indépendants (i.e., peer-reviewers) évaluant de manière critique (1). Le système de peer-review est considéré comme la meilleure méthode d'évaluation en recherche biomédicale (2). Il existe deux types de peer-review : l'évaluation de la demande de bourse ou subvention de recherche (i.e., *grant*) (3) et l'évaluation éditoriale. Au cours de notre travail, nous nous sommes uniquement intéressés à la peer-review éditoriale et par souci de concision nous utiliserons le terme anglais de peer-review.

### 1.1.2 Historique

L'émergence de l'imprimerie grâce à Gutenberg (1400-1468) a initié la possibilité de critique des écrits et donc en quelque sorte celui de la peer-review. En effet, ce qui était imprimé pouvait maintenant être distribué et affecter des citoyens dont la vie était à l'époque rythmée par le dogmatisme religieux.

Pour observer des avancées nouvelles sur le principe de relecture et de dissémination des idées, il faut attendre les travaux du philosophe Francis Bacon (1561-1626) sur l'intérêt d'une méthode universelle dans la génération, et l'évaluation des travaux scientifiques. Ce travail publié dans l'ouvrage *Novum Organum* (1620), inspira de nombreux érudits qui se livrèrent alors à des réunions secrètes pour débattre des moyens à mettre en œuvre pour garantir un développement de qualité de la science. Le 28 novembre 1660, un de ses groupuscules décida de sortir de l'ombre est de créer officiellement une société savante qui deviendra en 1662 *the Royal Society of London for Improving Natural Knowledge*. En 1665, la Société publia sa propre revue : *Philosophical Transactions.* C'est en 1752, que celle-ci se dota de ce que l'on considère comme les origines de la peer-review. En effet, les documents envoyés à la Société pour publication,

étaient soumis à l'inspection d'un groupe restreint de membres internes à la Société, qui connaissaient bien ces questions, et dont la recommandation à l'éditeur influait sur le devenir de ses manuscrits (4). Ce procédé fut par la suite adopté par de nombreuses sociétés (5). La problématique première de ce système était le délai de relecture du manuscrit. En effet, celui-ci devait passer entre les mains de plusieurs relecteurs et les moyens de communication de l'époque étaient rudimentaires. Il était donc convenu, à l'époque, que le contenu d'une recherche pouvait être rapporté par un membre titulaire de la Société, et ce, dans le but d'émettre rapidement les premières critiques et commentaires avant l'envoi officiel du manuscrit à la Société. Ce procédé nouveau dut rapidement faire face à des critiques et se retrouva confronté en premier lieu à l'écueil du conflit d'intérêt. En effet, la grande majorité des membres des Sociétés étaient des leaders d'opinion dans leur domaine. Ainsi, la présentation de certains travaux pouvait avoir un accueil mitigé. De ce fait, quand Charles Darwin (1809-1882) dut présenter les travaux d'Alfred Russel Wallace (1823-1913), un anthropologue travaillant lui aussi sur l'évolution des espèces, il s'est empressé de publier ses propres travaux pour le devancer (6). L'émergence des comités éditoriaux permit de limiter ce type de déconvenue.

Une des histoires les plus célèbres sur les limites de la peer-review est celle de l'épidémiologiste anglais John Snow (1813-1858) (7) lors de sa demande de financement pour la recherche sur le choléra. En effet, face à l'épidémie de choléra qui sévissait à Londres, celui-ci rédigea une demande de financement en émettant l'hypothèse que la transmission se faisait par l'eau. Il refusait donc de ce fait d'adhérer au dogme de l'époque selon laquelle les miasmes étaient le vecteur de la maladie. Face à la remise en cause d'un dogme aussi ancré dans l'esprit scientifique, le peer-review critiqua de manière très négative le projet de John Snow et celui-ci ne put obtenir de bourse.

Malgré cela, pour confirmer son hypothèse, John Snow réalisa en 1853 une enquête épidémiologique d'envergure en relevant chaque décès survenu en lien avec le choléra. Puis, identifiant le quartier de Soho comme probable épicentre de l'épidémie, John Snow mena une enquête auprès des habitants du quartier. Il trouva d'abord une relation significative entre le

nombre de cas de choléra dans un territoire, et la compagnie distributrice d'eau qui le dessert. Il contacta donc les autorités qui font enlever le bras de la pompe. De cette action s'en suivit une diminution significative du nombre de cas de choléra. Il soumit alors ses résultats au journal *The Lancet* qui refusa de publier son manuscrit du fait de défauts dans la méthodologie, et d'un manque d'argumentaire (8). Ces travaux furent reconnus comme significatifs par la même revue en 1855.

L'idée d'une peer-review externalisée n'est apparue qu'au milieu du XXéme siècle. La croissance massive des productions scientifiques, associée à leur spécialisation de plus en plus poussée, a obligé les revues à chercher des relecteurs en dehors de leur propre structure. Par exemple, le *Journal of American Medical Association* n'a utilisé une peer-review externe, qu'à partir de 1940 (9). A son origine, dans la grande majorité des processus de relecture les auteurs, ne savaient pas qui étaient leurs relecteurs. Ce n'est qu'à partir des années 1990 que l'on a observé l'émergence d'une peer-review dite en ouvert (i.e., open peer-review), où l'anonymat des auteurs ou des correcteurs ou des 2, peut être levé (10) (11).

En 1989, Drummond Rennie, un éditeur du *Journal of American Medical Association*, décida d'organiser le premier congrès consacré à la peer-review (10). En effet, pour Drummond Rennie, il était primordial d'améliorer la peer-review en la rendant plus scientifique (12). Suite à ce congrès, la recherche dans ce domaine s'est développée, dans le but d'étudier l'ensemble des mécanismes de la peer-review et donc d'identifier des leviers d'améliorations (11) (14). Depuis, le congrès se tient tous les 4 ans, et a pour objectifs d'échanger entre les différents acteurs de cette recherche.

### 1.1.3 Les objectifs de la peer-review

Selon les éditeurs, la peer-review a deux objectifs bien distincts : aider les éditeurs à la prise de décision sur le devenir d'un manuscrit, et améliorer la qualité du manuscrit(15). Or ces deux objectifs ne sont pas toujours explicites. De plus, les facteurs importants associés à la prise de décision varient en fonction du journal. Par exemple, le journal *New England Journal of Medicine* semble s'attacher plus particulièrement à l'originalité de la question de recherche, alors que le journal *PlosOne* s'attache principalement à la méthodologie de l'article. Le manque de clarté dans les objectifs de la peer-review peut être une des causes de l'impact limité de la peer-review sur la qualité des articles publiés.

### 1.1.4 Le processus de peer-review

Le processus de la peer-review suit classiquement des étapes clairement définies (Figure 1). La soumission du manuscrit au journal constitue la première étape et est réalisée par l'auteur correspondant via une plateforme éditoriale dédiée (e.g., Scholar-One Manuscript). Le papier est alors évalué une première fois par l'office éditorial dont l'objectif est d'évaluer le format du manuscrit (par exemple, le respect des recommandations aux auteurs). A ce stade, la qualité du manuscrit n'est pas encore évaluée. Le manuscrit est alors transmis à l'éditeur en chef ou un des éditeurs associés qui évaluent si le manuscrit correspond à la thématique du journal, et s'il est suffisamment original pour être envoyé en relecture aux peer-reviewers. Dans la négative, le papier est rejeté sans être évalué par les peer-reviewers. Certains journaux assignent, à cette étape, un éditeur associé qui va gérer le processus de peer-review. Mais cette étape n'est pas généralisée. L'éditeur en charge du processus de peer-review (i.e., éditeur en chef ou associé) envoie des invitations à des peer-reviewers dont il pense que l'expertise permettra une évaluation de qualité du manuscrit. Le nombre de peer-reviewers dépend du type de manuscrit et du journal.

Il est communément accepté qu'il faut au moins 2 peer-reviewers par évaluation. Les peer-reviewers sélectionnés par l'éditeur peuvent décliner son invitation et ce, pour de multiples raisons : défaut de temps, expertise insuffisante dans le domaine concerné, conflit d'intérêt… Dans ce cas, le peer-reviewer peut proposer un remplaçant. Sinon, l'éditeur se doit de trouver un autre peer-reviewer. Une fois les peer-reviewers identifiés, ceux-ci vont produire un rapport d'évaluation sur le manuscrit soumis. Avant cette évaluation, les peer-reviewers reçoivent des instructions de la part de l'éditeur sur comment diriger son évaluation. C'est à ce stade que la qualité du manuscrit est évaluée. Une fois l'évaluation faite, les peer-reviewers la renvoient à l'éditeur avec une recommandation quant au devenir du manuscrit : acceptation sans révision, acceptation avec révision(s) (mineure(s) ou majeure(s)) ou rejet. Une fois le retour de tous les peer-reviewers fait, l'éditeur prend connaissance des remarques et des recommandations dans l'objectif de la prise de décision définitive quant au devenir du manuscrit. L'éditeur a la main sur la décision finale. Il peut décider de suivre ou non les recommandations des peer-reviewers ou bien demander un avis supplémentaire à un autre peer-reviewer. Une fois la décision prise par l'éditeur, celui-ci la communique à l'auteur correspondant par courriel en incluant les remarques des peer-reviewers ainsi que son propre commentaire. Si le papier est accepté, celui-ci part en production. Dans le cas de révisions, l'auteur devra produire une nouvelle version en tenant compte des remarques et dans un délai imparti par l'éditeur.

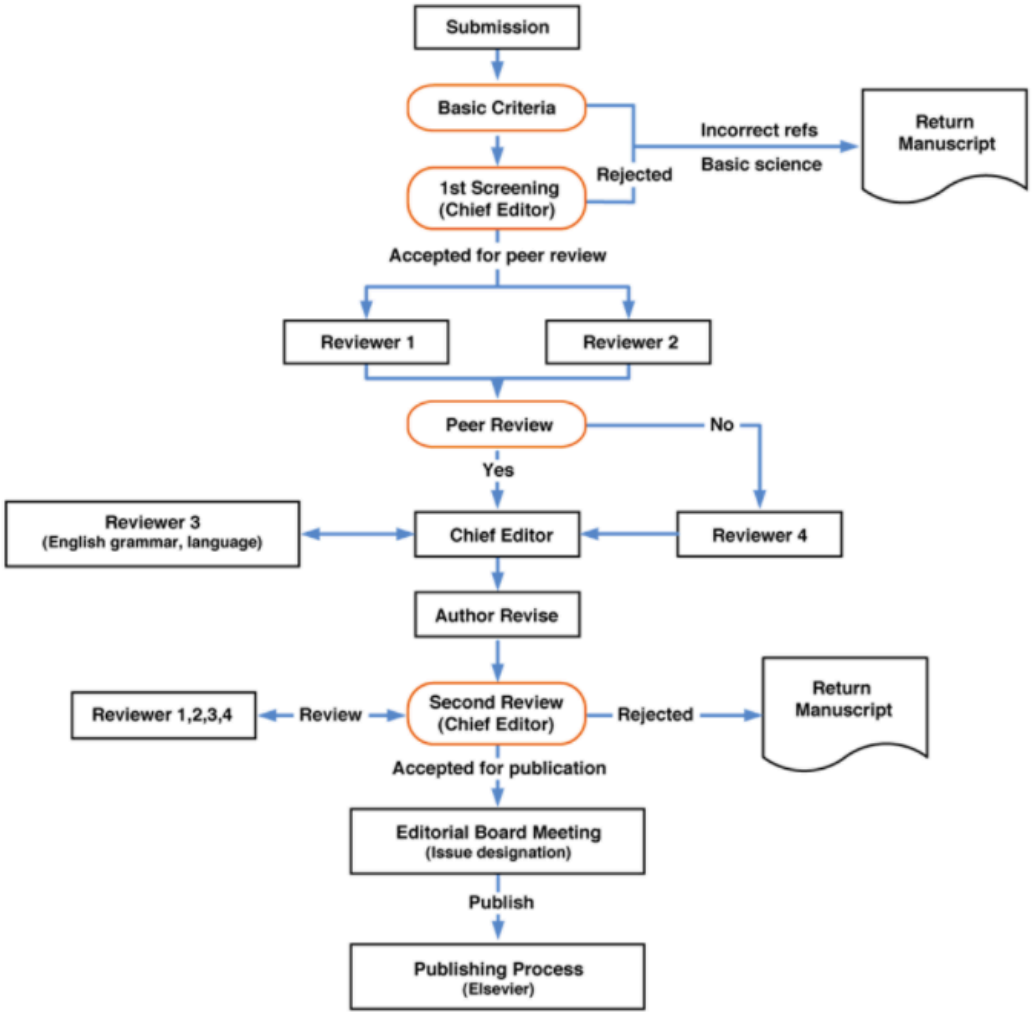



**Figure 1. Les différentes étapes du processus de la peer-review (infographie d'*Annals of Emergency Medicine*)**

### 1.1.5 Les différentes modalités de la peer-review

#### 1.1.5.1 L'évaluation pré publication

Comme explicité ci-dessus, la forme la plus rependue d'évaluation des manuscrits se fait avant que ceux-ci ne soient publiés. C'est ce que l'on appelle l'évaluation pré publication.

Ce type d'évaluation peut se faire selon différentes modalités :

1) Evaluation en simple aveugle : c'est la méthode d'évaluation la plus répandue. Les noms des peer-reviewers sont cachés aux auteurs. L'avantage de cette méthode est que les peer-reviewers bénéficient d'une liberté leur permettant de ne pas craindre des représailles de la part des auteurs.

En revanche, l'inconvénient de cette méthode est que les peer-reviewers peuvent de sentir déresponsabilisés et donc faire des commentaires agressifs voir déplacés.

2) Evaluation en double aveugle : ni l'identité des auteurs, ni celle des peer-reviewers ne sont communiquées. Les avantages de cette méthode sont qu'elle évite le biais de relecture (i.e., favoriser la publication d'un pays ou d'une institution), ou encore favoriser la publication d'auteurs de renoms. Bien entendu, cet aveugle peut être brisé si les peer-reviewers identifient les auteurs du fait du style d'écriture ou encore par les citations utilisées par les auteurs. De plus, cette modalité de peer-review est complexe à organiser pour les éditeurs.

3) Evaluation en ouvert : l'identité des auteurs et des peer-reviewers sont révélées. Cette méthode a autant de détracteurs que de supporters. Pour certains, cette méthode est préférable car elle limiterait le nombre de commentaires mal venus, le risque de plagiat et encouragerait une évaluation honnête. Pour d'autres, cette méthode est contre-productive car elle oblige les peer-reviewers à modérer leurs critiques. De plus, le fait pour un peer-reviewer de devoir évaluer un article d'un auteur influent, pourrait intimider celui-ci et donc atténuer ses éventuelles critiques.

#### 1.1.5.2 L'évaluation post publication

L'évaluation post publication (i.e., post-publication peer-review) est le fait d'évaluer et de commenter les manuscrits une fois qu'ils sont publiés, les commentaires étant ouverts à tous. Le premier article prenant position sur l'intérêt de cette méthode a été publié en 2007 dans *Nature* (16). Face à la critique grandissante de la peer-review classique, de nombreux blogs ou plateformes participatives ont vu le jour.

A ce jour, il existe trois modèles distincts de peer-review post publication. Tout d'abord, la relecture par des peer-reviewers invités par des éditeurs pour évaluer des articles publiés en ligne et n'ayant pas encore été évalués. Ce système est celui utilisé par F1000Research (17). Les articles sont donc publiés en ligne mais non indexés dans les bases de données (i.e., PubMed) et ce jusqu'à la relecture positive de deux peer-reviewers. La dernière version du manuscrit sera donc publiée sur le site de l'éditeur et indexé dans les base de données. L'ensemble des versions

et des critiques des peer-reviewers sont consultables.

La critique d'articles publiés et indexé sur des blogs spécifiques (i.e., PubPeer) (18) par des tiers est le mode de peer-review post publication le plus répandu actuellement. Ce type d'échange peut permettre de mettre en évidence des problèmes méthodologiques non mis en évidence lors de la peer-review classique.

Enfin, l'évaluation par des peer-reviewers volontaires d'article publié et indexé est actuellement testé. La plateforme Science Open (19) propose depuis 2013 aux auteurs de mettre en ligne leurs travaux pour promouvoir leur recherche mais pour pouvoir aussi bénéficier de critique de la part de peer-reviewers volontaires. Les peer-reviewers volontaires doivent par contre avoir une qualification (i.e., 5 articles publiés et indexés).

L'impact de la relecture post publication est difficilement évaluable (20). L'ensemble de la communauté scientifique semble d'accord sur l'intérêt d'un tel processus (21) (22) mais celle-ci doit être encadrée et structurée (23) (24) (25).

### 1.1.6 Les critiques de la peer-review

#### 1.1.6.1 Coût et délai

En 2010, un rapport commandé par la Chambre des communes du Royaume-Uni montrait que le coût, pour les établissements d'enseignement supérieur britanniques en termes de temps de travail, était de 125 à 189 millions d'Euros par an pour le processus de peer-review, et jusqu'à 34 millions d'Euros par an pour le travail éditorial. Dans le monde entier, on estime que l'évaluation par les peer-reviewers coûte environ 2,2 milliards d'Euros par an et représente environ un quart des coûts globaux de la publication scientifique (26). Les ressources humaines ont été estimées à environ 15 millions d'heures en 2010 (27).

#### 1.1.6.2 Reproductibilité

De nombreux essais ont mis en évidence le manque de reproductibilité entre les peer-reviewers. La reproductibilité a été étudié à différents niveaux : la cotation de résumé, la qualité de la peer-review ou encore la recommandation faite aux éditeurs sur l'acceptation ou non du manuscrit. En 2006, Rowe et al. (28) se sont intéressés à l'accord entre les peer-reviewers sur le score des résumés soumis au congrès annuel de médecine d'urgence Canadienne. L'analyse à inclus au final 495 résumés soumis durant une période de 4 ans. Les auteurs ont calculé le coefficient de corrélation intra classe pour des items spécifiques (i.e., méthodologies, analyse statistiques, présentation…) ainsi que pour le score global. En 4 ans, le coefficient de corrélation intra classe était passé de faible (0,36 ; 95% IC : 0,24-0,49) à modéré (0,59 ; 95% IC : 0,50-0,68). Les peer-reviewers était moins d'accord sur les items individuels sur l'ensemble de la période. Par exemple, sur l'évaluation du design de l'étude, le coefficient de corrélation intra classe était passé de 0,15 (95% IC : 0,03-0,29) à 0,23 (95% IC : 0,13-0,33). Les auteurs avaient avancé comme hypothèse que l'absence de critère objectifs quant à l'évaluation par les peer-reviewers pouvait être à l'origine de cette absence d'accord.

L'équipe éditoriale du *Journal of General Internal Medicine* (29) a évalué le score global du manuscrit suite à l'évaluation par les peer-reviewers et son impact sur les décisions éditoriales. Ils ont certes mis en évidence que le score global moyen des articles accepté était supérieur à celui des articles rejetés (1,6 versus 3,4). Mais les auteurs ont mis en évidence qu'il n'y avait qu'une corrélation faible entre scores donnés par les peer-reviewers. En effet, le coefficient de corrélation intra classe allait de 0,09 à 0,13 selon si l'article était rejeté, révisé ou accepté.

Un journal orthopédique, *The spine journal* (30), s'est focalisé sur la reproductibilité quant aux recommandations faites aux éditeurs sur l'acceptation ou non des manuscrits. Les auteurs ont été randomisé en 2 groupes distincts : un groupe article accepté/accepté avec révisions mineures et un groupe accepté avec révisions majeures/rejeté. Deux cents articles consécutifs ont été inclus

dans chaque groupe. Le coefficient de corrélation intra classe était respectivement faible à 0,155 (95% IC : 0,017-0,294) et faible à 0,07. Par ces résultats les auteurs mettaient en valeur la variabilité d'examen entre les peer-reviewers et qu'ainsi, il faudrait travailler sur l'étude d'éléments visant à standardiser l'examen par les peer-reviewers.

#### 1.1.6.3 Présentation inadéquate des manuscrits

Bien qu'une des missions premières de la peer-review soit la détection des erreurs méthodologiques dans les manuscrits, il semblerait que les peer-reviewers rencontrent des difficultés dans la réalisation de cette tâche.

En effet, de nombreuses études ont mis en évidence que le système actuel peut ne pas identifier toutes les failles dans la qualité des manuscrits soumis et des articles publiés (34) (35). Une étude de 2014, conduite par Hopewell et al. (36), a mis en évidence que les peer-reviewers ne parvenaient pas à détecter d'importantes déficiences dans l'évaluation de la qualité de la présentation (i.e., »reporting »). Les auteurs ont réalisé une étude avant/après l'évaluation des peer-reviewers. Sur les 93 essais randomisés inclus dans leur analyse, 38% (n = 35) ne décrivaient pas la méthode de génération de randomisation, 54% (n = 50) la méthode d'allocation, 50% (n = 46) si l'étude était en aveugle, 34% (n = 32) le calcul de la taille de l'échantillon, 35% (n = 33) n'explicitait pas le critère de jugement principal et secondaire, 55% (n = 51) les résultats pour le critère de jugement principal, et 90% (n = 84) les détails du protocole de l'essai.

Les peer-reviewers semblent aussi avoir des difficultés à détecter les défauts dans les analyses statistiques. Une équipe Française s'est ainsi intéressé à la qualité des éléments statistiques dans les essais contrôlés randomisés de non-infériorité ou d'équivalence (37). Sur la période d'analyse de 2 ans, 162 articles ont été inclus dans l'analyse. Les auteurs ont mis en évidence qu'un quart (n=35, 22%) des articles ne rapportaient pas tous les éléments nécessaires au calcul du nombre de sujet nécessaire. De plus, seulement 33 (20%) des articles explicitaient une justification quant à la borne de non-infériorité.

De nombreuses revues systématiques d'essai clinique ont mis en évidence ce manque de détection de défaut d'éléments clés par les peer-reviewers (38)(39)(40)(41)(42).

#### 1.1.6.4 Détection des spins

On appelle spin une distorsion des résultats d'une étude, qu'elle soit intentionnelle ou non. Ce manquement a 2 conséquences directes : 1) le risque de mauvaise interprétation des essais par les cliniciens (31) pouvant faire courir un risque au patient, et 2) le gaspillage de la recherche.

Lazarus et al. (32) ont publié en 2014 une revue systématique et une étude avant-après pour évaluer l'impact des peer-reviewers sur les spins dans les résumés des essais non-randomisés évaluant une intervention thérapeutique. Sur les 128 articles analysés, les peer-reviewers ont identifiés environ un quart des spins (n=31, 24%) présents dans la conclusion du résumé dans 97 (76%) articles. De plus, les peer-reviewers ont proposé de rajouter des spins dans 19 (15%) articles (Figure 2).

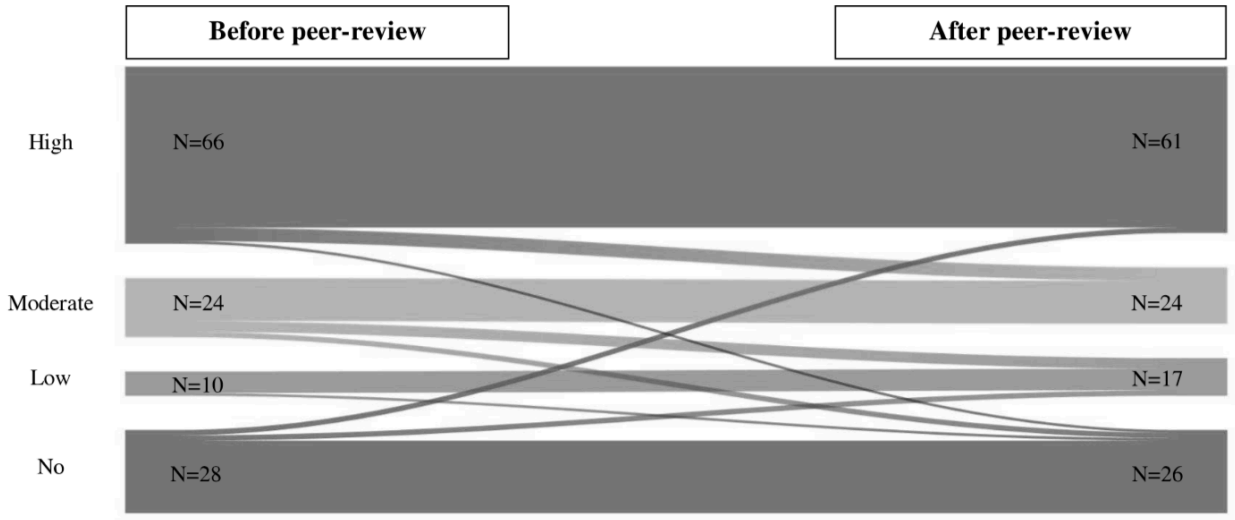



**Figure 2. Niveau des spins dans la conclusion des résumés, avant et après peer-review**

La difficulté pour les peer-reviewers de détecter des spins a aussi été mis en évidence dans une revue systématique d'essais évaluant une intervention précoce dans les psychoses (33). En effet, les auteurs ont analysé la qualité de 38 essais contrôlés randomisés ou quasi expérimentaux

publié entre 2000 et 2013. Parmi ces articles, 22 (66%) contenaient des spins. Le spin le plus fréquent était le changement de critère de jugement principal et une fixation des résultats sur le critère de jugement secondaire.

La présence de spin dans les résumés a un impact sur l'interprétation des résultats (31).

#### 1.1.6.5 Changement du critère de jugement

Certains items ne demandent pas aux peer-reviewers une expertise spécifique (43). On peut citer par exemple l'évaluation de la concordance des critères de jugement principaux et secondaires entre le registre ou le protocole et l'essai publié. Or cet item semble poser des difficultés aux peer-reviewers. En effet, Mathieu et al. (44) ont publié en 2009 un article ayant comparé le critère de jugement principal rapporté par les auteurs avec celui pré spécifié dans le registre. Sur les 323 essais inclus, environ la moitié (n=147, 46%) étaient correctement enregistrés (i.e., enregistrement avant la fin de l'essai et le critère de jugement principal clairement pré spécifié). Parmi ceux-ci 46 (31%) avaient des divergences entre le critère de jugement pré spécifié et celui rapporté par les auteurs dans la publication. De plus, un projet récent COMPare-Trials.org (45), a systématiquement évalué, entre octobre 2015 et janvier 2016, les critères de jugement pré spécifiés dans le protocole et/ou l'enregistrement de l'étude, et ceux rapportés dans l'article. Sur 67 essais évalués, 354 critères de jugements n'étaient pas rapportés, et 357 ont été ajoutés.

Ce défaut dans la détection des changements de critère de jugement a été démontré dans différentes spécialités : neurologie (46), psychiatrie (47), chirurgie cardio-vasculaire (48) etc…

Une étude empirique a montré que les chances de rejet des revues de haut rang n'étaient pas très différentes pour les essais avec (79%) ou sans (71%) écarts de résultats primaires (49).

#### 1.1.6.6 Détection des erreurs

Les peer-reviewers semblent aussi en difficulté pour mettre en évidence les erreurs présentes

dans les manuscrits soumis à leur évaluation. En effet, l'équipe éditoriale du BMJ a réalisé un essai contrôlé randomisé avec 607 peer-reviewers pour évaluer si le fait de les entrainer leur permettait de détecter des erreurs méthodologiques (50). Pour ce faire, les peer-reviewers ont été randomisés en 3 bras : entrainement en présentiel, entrainement seul avec un programme standardisé et pas d'entrainement. Chaque peer-reveiwers évaluaient les 3 mêmes articles, chaque article ayant 9 erreurs. Le nombre moyen d'erreurs détectées dans chaque article était faible à savoir : 2,58, 2,71 et 3,0. Il n'y avait pas de différences entre chaque bras. Cette étude met en évidence la difficulté des peer-reviewers à détecter des erreurs et ce malgré un entrainement.

#### 1.1.6.7 Détection des conflits d'intérêt

Le débat sur la détection des conflits d'intérêts par les peer-reviewers est principalement dû au fait que des articles publiés semblent avoir été biaisés par les intérêts commerciaux ou autres.

Ces dernières années, on a eu l'impression que les conflits d'intérêts avaient un effet néfaste sur la science médicale puisqu'ils influençaient la politique de santé et la pratique clinique, ce qui a poussé les revues médicales à contredire ou à compenser le problème (51) (52) (53). Une des études majeures sur les conflits d'intérêt financiers dans les essais cliniques randomisés rapportés dans le BMJ de 1997 à 2001 (159 articles de 12 spécialités) a révélé que les conclusions des auteurs étaient significativement plus positives pour les essais financés par des organisations à but lucratif ($p = 0,014$)(54).

Dans l'un de ses principaux documents publié en 2009, l'Association mondiale des rédacteurs médicaux « *World Association of Medical Editors »* (55) recommandait de signaler les conflits d'intérêt des auteurs et des rédacteurs en chef liés aux soumissions des auteurs (56). Dans ses directives à l'intention des éditeurs, le Comité d'éthique des publications « *Committee on Publication Ethics* » a conseillé de mettre en place des systèmes de gestion des conflits d'intérêts, non seulement des auteurs et des rédacteurs mais aussi des peer-reviewers (57). Enfin, ces diverses associations ont récemment souligné la nécessité de réviser régulièrement les

consignes des revues, et d'adopter des procédures pour rendre publique l'ensemble des conflits d'intérêts. Mais il semblerait que les éditeurs ne prennent pas la dimension des conflits d'intérêts des peer-reviewers en compte (58). En effet, une enquête sur la politique éditoriale des 37 journaux anglophones d'ophtalmologie a révélé que seuls 60% (n=18) avaient une politique claire quant aux conflits d'intérêts pour les peer-reviewers. En 2004, une enquête auprès des éditeurs de 30 revues de médecine générale a mis en évidence que seules 9 revues (30%) avaient établi une politique claire de déclaration des conflits d'intérêts des éditeurs. Douze journaux (37%), n'avaient pas l'intention de déclarer les conflits d'intérêts financier des éditeurs.

L'ensemble des acteurs de la recherche (i.e., auteurs, peer-reviewers et éditeurs) doivent être des cautions morales et doivent donc s'efforcer à mettre en évidence les informations sur le financement des essais, le soutien logistique, l'implication des sponsors dans la recherche. La mise en évidence de ces données est à présent fondamentale, et fait l'objet d'une section particulière dans les checklists de soumission des manuscrits aux éditeurs (59).

La non prise en compte des conflits d'intérêt a été mise en évidence dans de nombreux autres domaines de la recherche biomédicale, en particulier dans les revues systématiques et les méta-analyses (60) (61). Par exemple, sur les 151 revues systématiques publiées dans la base de données Cochrane en 2010, seulement 46 (30%) ont fourni des déclarations sur les sources de financement des essais contrôlés et 16 (11%) sur les liens entre l'auteur et l'industrie (61). Ce manque de transparence peut conduire à un biais de publication dans le cas potentiel où le ou les peer-reviewers ont un conflit d'intérêt avec le laboratoire sponsorisant l'étude.

#### 1.1.6.8 Les biais cognitifs

##### 1.1.6.8.1 Biais d'objectivité

L'objectivité des éditeurs et des peer-reviewers peut également être prise à défaut par des préjugés, des croyances ou hypothèses (63)(62).

Il a également été suggéré que les éditeurs ou les peer-reviewers évaluent défavorablement les manuscrits contenant des idées non conventionnelles. Deux exemples souvent cités sont la description de Hans Krebs du cycle de l'acide citrique, et la description de Barbara McClintock des éléments du génome, qui ont tous deux été rejetés par Nature (63). Les deux auteurs ont plus tard remporté le prix Nobel respectivement en 1953 et 1983, pour leurs découvertes. Ces deux exemples mettent en évidence que les croyances des peer-reviewers et des éditeurs pourraient ralentir l'avancée de la science.

##### 1.1.6.8.2 Biais de genre

La présence d'un biais de genre parmi les évaluateurs a également été débattu (66). En 1990, Lloyd et al. (65) a soumis à l'évaluation par des peer-reviewers 2 articles pré fabriqués dont l'un avait des auteurs masculins, et l'autres, des auteurs féminins. Les manuscrits d'auteur féminin ont été acceptés significativement, plus souvent, par les femmes évaluatrices (62%) que par les évaluateurs masculins (21%). Les peer-reviewers masculins n'ont pas discriminé entre des manuscrits d'auteurs masculins ou féminins en termes du taux d'acceptation (20-30%). En revanche, les peer-reviewers féminins étaient également beaucoup moins susceptibles d'accepter la version masculine d'un papier (10%) par rapport à la version féminine (62%). Gilbert et al. (66) ont évalué les préjugés sexistes au *Journal of American Medical Association* dans une rétrospective étude de cohorte de 1 851 articles. Les auteurs ont analysé les pratiques des peer-reviewers et des éditeurs selon leur genre. Ils ont conclu au fait qu'il n'y avait pas effet du genre sur la recommandation finale de la publication, ou, sur l'acceptation finale à la publication. En revanche, les éditeurs féminins avaient des charges de travail plus importantes et étaient plus susceptibles de rejeter les manuscrits rapidement. En examinant le sexisme dans le processus

d'évaluation par les peer-reviewers appliqué au financement, Hargens et al. (67) a mis en évidence que les subventions étaient accordées à des hommes et à des femmes à un taux identique. Mais il est difficile de conclure sur cette seule étude observationnelle. Une revue systématique récente a évalué les interventions testées pour prévenir, ou diminuer, le biais de genre dans la demande de financement. Une seule étude a été retrouvée, ce qui amène à suggérer qu'il serait intéressant d'évaluer ce biais par une étude interventionnelle (68).

#### 1.1.6.9 Conduites inappropriées

Le comité international des éditeurs des journaux médicaux (i.e., *International Commitee of Medical Journal Editors*) (69) a publié des règles de bonnes conduites destinées aux peer-reviewers. Ce comité rappelle que tous articles soumis à l'évaluation des peer-reviewers sont confidentiels et que de ce fait le peer-reviewer ne peut en aucun cas discuter avec un tiers de ces travaux et encore moins s'en inspirer. Par ailleurs, les peer-reviewers doivent se récuser s'ils considèrent avoir un conflit d'intérêt avec le travail soumis à évaluation. Malgré ces directives claires, certaines mauvaises pratiques ont été mises en évidence.

##### 1.1.6.9.1 Plagiat

Il y a plusieurs façons d'abuser du processus de la peer-review. En effet, les peer-reviewers peuvent voler des idées et les présenter comme les leurs, ou produire une critique injustement sévère pour bloquer ou au moins ralentir la publication des idées d'un concurrent. Dr Drummond Rennie raconte l'histoire d'un article qu'il a envoyé pour relecture a une collaboratrice lorsqu'il était rédacteur en chef adjoint du *New England Journal of Medicine* (70). Celle-ci ayant produit une critique du document, a copié certains des paragraphes et les a soumis à un autre journal, *l'American Journal of Medicine*. Ce journal, par coïncidence, l'a envoyé pour examen au patron de l'auteur du papier plagié qui a donc compris qu'il avait été plagié. Récemment, un article a dû être rétracté par l'auteur car celui-ci s'était accaparé les données d'un manuscrit qu'il lui avait

été demandé de relire et critiquer (71) (72). Ce papier correspondait à un travail de thèse ayant pris près de cinq ans à être réalisé pour l'étudiant (73).

#### 1.1.6.9.2 Fausse peer-review

Des cas de fausses peer-review ont été mis en évidence. Lors de la soumission d'un manuscrit, l'auteur doit généralement proposer des noms de peer-reviewers possibles. Ce type de fraude, correspond au fait que l'auteur soumet des noms d'experts avec une fausse adresse mail qui correspond à un compte de l'auteur. Cette pratique mise en évidence récemment a débouché sur la rétraction de nombreux papiers par les éditeurs dans de nombreux domaines (74)(75)(76)(77)(79).

## 2. Objectifs de la thèse

Les objectifs de la thèse étaient :

1) d'explorer le processus de peer-review en évaluant les attentes des éditeurs et des peer-reviewers.

2) d'identifier les interventions visant à améliorer la peer-review et ayant démontré leur efficacité.

3) de développer et évaluer l'impact d'un outil informatisé permettant de standardiser les critiques des reviewers sur la qualité de la présentation des articles.

# 3. Les tâches incombant aux peer-reviewers lors de l'évaluation d'un essai contrôlé randomisé

## 3.1 Justification de l'étude

L'évaluation d'un article scientifique par des peer-reviewers implique différentes tâches. Le peer-reviewer peut par exemple évaluer la pertinence de la question, la qualité méthodologique, la qualité de la présentation ou encore l'applicabilité des résultats. Le nombre de tâches à réaliser lors d'une peer-review est très important. Or les éditeurs et les peer-reviewers peuvent avoir une appréciation différente de la difficulté et de l'importance de chaque tâche. De plus, ces tâches ne sont pas toujours clairement exprimées par les éditeurs dans leurs recommandations et celles-ci peuvent varier d'un éditeur à l'autre. Enfin, les peer-reviewers qui font des reviews pour de nombreux journaux ne lisent pas toujours les recommandations spécifiques du journal.

Notre hypothèse est que trop de tâches sont attendues de la part des peer-reviewers et que ceux-ci ont une vision différente, de celle des éditeurs, sur quelles tâches sont les plus importantes.

## 3.2 Objectifs

Les objectifs de cet article étaient : a) d'identifier toutes les tâches attendues des peer-reviewers lors de l'évaluation d'un manuscrit rapportant un essai contrôlé randomisé, b) de classer ces tâches en fonction de leur importance et c) de déterminer quelles tâches sont clairement demandées par les éditeurs dans leurs recommandations aux peer-reviewers.

## 3.3 Méthodes

Nous avons procédé en 3 étapes :

- Identification des tâches attendues des peer-reviewers lors de l'évaluation d'un manuscrit rapportant un essai contrôlé randomisé.

Dans un premier temps, nous avons réalisé une revue systématique de la littérature biomédicale dans les bases de données Medline via Pubmed, EMBASE et Google Scholar tous les articles traitant de la peer-review et des tâches incombant aux peer-reviewers. Cette recherche s'est intéressée aux articles publiés en anglais sur les dix dernières années. Ensuite, nous avons analysé les recommandations faites aux peer-reviewers par des journaux biomédicaux. A savoir : a) les journaux ayant publié au moins 15 essais contrôlés randomisés sur une période de 3 mois, b) les 10 journaux avec un impact factor le plus élevé parmi 14 spécialités médicales. Une fois les journaux identifiés, nous avons contacté les éditeurs pour qu'ils nous transmettent les recommandations qu'ils envoyaient spécifiquement aux peer-reviewers. Nous avions aussi analysé les recommandations aux peer-reviewers sur le site internet de l'éditeur.

Notre revue systématique nous a permis d'identifier 2 117 références. Après lecture, 90 articles ont été inclus dans l'analyse finale. Notre échantillon de journaux biomédicaux comportait 171 journaux.

A partir de ces différentes sources, nous avons relevé l'ensemble des tâches demandées aux peer-reviewers lors de l'évaluation d'un essai contrôlé randomisé. Enfin, nous avons regroupé les tâches proches.

- Enquête auprès de peer-reviewers d'essais contrôlés randomisés pour classer les tâches en fonction de leur importance.

Nous avons invité à participer à l'étude : des peer-reviewers de 3 journaux (79), les éditeurs des journaux identifiés lors de l'étape précédente, les auteurs correspondants des 10 journaux ayant le plus haut impact factor dans le domaine *Public, Environmental & Occupational Health* ainsi que dans les revues: *Statistics in Medicine, Biostatistics, Biometrics,* et *Statistical Methods in Medical Research* et les auteurs correspondants des essais randomisés publiés et indexés PubMed entre le 1er janvier 2014 et le 15 juin 2014.

Chaque participant devait classer les tâches que nous avions condensées en fonction de leur importance, et ce, selon la méthode Q-Sort (80) (81). Le score allait de -5, tâche la moins importante, à +5, tâche la plus importante.

- Evaluation des tâches clairement demandées aux peer-reviewers par les éditeurs de journaux biomédicaux.

Pour chaque tâche condensée, nous avons recherché dans la rubrique dédiée aux peer-reviewers, sur le site internet du journal, ainsi que les documents transmis par les éditeurs, si celles-ci étaient clairement demandées.

## 3.4 Résultats

- Identification des tâches attendues des peer-reviewers lors de l'évaluation d'un manuscrit rapportant un essai contrôlé randomisé.

Nous avons pu identifier 205 tâches que nous avons condensé en 36 tâches principales.

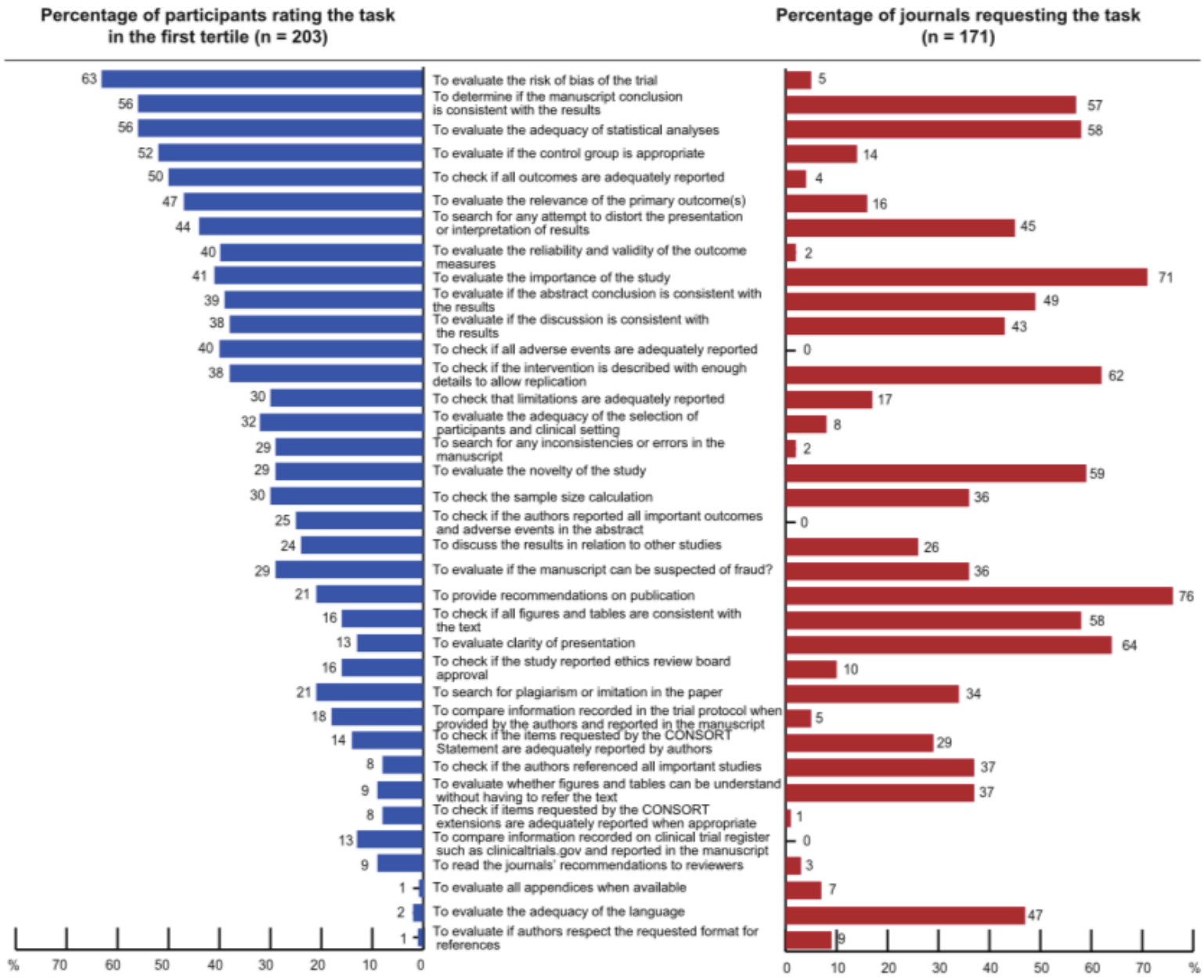
- Enquête auprès de peer-reviewers d'essais contrôlés randomisés pour classer les tâches en fonction de leur importance.

Pour les 229 peer-reviewers ayant complété l'étude, les tâches les plus importantes sont celles se rapportant à la méthodologie (i.e., évaluer le risque de biais, déterminer si la conclusion du manuscrit est congruente avec les résultats, évaluer l'adéquation de l'analyse statistiques).

- Evaluation des tâches clairement demandées aux peer-reviewers par les éditeurs de journaux biomédicaux.

Les tâches classées dans le premier tertile d'importance (≥2 sur l'échelle de -5 à +5) n'étaient pas conformes aux tâches les plus fréquemment demandées par les éditeurs. Par exemple, l'évaluation du risque de biais a été notée dans le premier tertile par 63% des participants mais clairement demandée par seulement 5% des éditeurs (Figure 3). En revanche, la tâche la plus fréquemment demandée par les éditeurs était la recommandation pour la publication (76%), mais cette tâche a été classée dans le premier tertile par seulement 21% de tous les participants.

**Figure 3. Proportion de participants évaluant la tâche dans le premier tertile (c.à.d ≥2) et proportion d'éditeurs explicitant la tâche dans leurs recommandations aux auteurs**



## 3.5 Discussion et limites

Cette étude montre que les peer-reviewers et les éditeurs n'ont pas la même vision des tâches importantes lors de l'évaluation d'un essai randomisé : les éditeurs faisaient surtout des recommandations sur le format (i.e., tableaux/figures/présentation) et sur l'importance de l'étude, alors que les peer-reviewers étaient surtout intéressés par la méthodologie. De plus, les recommandations actuelles des éditeurs peuvent ne pas être suffisamment explicites. Les recommandations des éditeurs étaient principalement génériques et non spécifiques de la conception des études. Cela est probablement lié au fait que les rédacteurs se fient généralement à l'expertise du peer-reviewer et s'attendent à ce qu'il sache quels aspects sont importants et quels aspects doivent être évalués. Cependant, certains éléments de preuve dans la littérature montrent que l'impact des peer-reviewers est discutable (36) (82) (83) (84). Par ailleurs, nous avons mis en évidence des différences dans le classement de l'importance des tâches selon l'expertise du peer-reviewer.

Des tâches importantes et simples, comme par exemple, comparer les items enregistrés dans le registre d'étude (i.e., *ClinicalTrials.gov*), qui peuvent facilement améliorer la qualité du manuscrit sont peu considérées par les peer-reviewers et peu demandées par les éditeurs. Or, cette simple vérification peut grandement limiter le risque de sélection du critère de jugement (44).

Notre étude présente des limites. Tout d'abord, un biais de recrutement lié au fait que nous ne savons pas si les peer-reviewers qui ont répondu sont représentatifs de l'ensemble des peer-reviewers. Cela est renforcé par un faible taux de réponse. De plus, seuls 45% des éditeurs des journaux inclus nous ont transmis leurs documents. Enfin, notre étude était une étude réalisée à un moment donné (juillet 2014), et entre le moment de recueil des données et celui de la publication, des mises à jour par les éditeurs ont pu être faites.

## 3.6 Conclusion

Les tâches considérées comme les plus importantes par les peer-reviewers (i.e., la méthodologie, les statistiques, et les résultats) ne sont pas fréquemment demandées par les éditeurs.

## 3.7Article

Chauvin et al. BMC Medicine
DOI 10.1186/s12916-015-0395-3

BMC Medicine

RESEARCH ARTICLE Open Access

# The most important tasks for peer reviewers evaluating a randomized controlled trial are not congruent with the tasks most often requested by journal editors

Anthony Chauvin[2,3†], Philippe Ravaud[1,2,3], Gabriel Baron[1,3], Caroline Barnes[2,3] and Isabelle Boutron[1,2,3*†]

**Abstract**

**Background:** The peer review process is a cornerstone of biomedical research publications. However, it may fail to allow the publication of high-quality articles. We aimed to identify and sort, according to their importance, all tasks that are expected from peer reviewers when evaluating a manuscript reporting the results of a randomized controlled trial (RCT) and to determine which of these tasks are clearly requested by editors in their recommendations to peer reviewers.

**Methods:** We identified the tasks expected of peer reviewers from 1) a systematic review of the published literature and 2) recommendations to peer reviewers for 171 journals (i.e., 10 journals with the highest impact factor for 14 different medical areas and all journals indexed in PubMed that published more than 15 RCTs over 3 months regardless of the medical area). Participants who had peer-reviewed at least one report of an RCT had to classify the importance of each task relative to other tasks using a Q-sort technique. Finally, we evaluated editors' recommendations to authors to determine which tasks were clearly requested by editors in their recommendations to peer reviewers.

**Results:** The Q-sort survey was completed by 203 participants, 93 (46 %) with clinical expertise, 72 (36 %) with methodological/statistical expertise, 17 (8 %) with expertise in both areas, and 21 (10 %) with other expertise. The task rated most important by participants (evaluating the risk of bias) was clearly requested by only 5 % of editors. In contrast, the task most frequently requested by editors (provide recommendations for publication), was rated in the first tertile only by 21 % of all participants.

**Conclusions:** The most important tasks for peer reviewers were not congruent with the tasks most often requested by journal editors in their guidelines to reviewers.

**Keywords:** Peer review, Q-sort, Randomized controlled trials, Recommendations to reviewers

## Background

The peer review process is a cornerstone of biomedical research publication [1]. The editorial peer review process is described as journal editors relying on the views of independent experts in making decisions on, for example, the publication of submitted manuscripts or presentation of reports at meetings [2]. The peer review system is considered the best method for evaluating publications of health research [3]. Nevertheless, the effectiveness of the peer review system is questioned. In 2010, a report ordered by the UK House of Commons showed that the cost to the UK Higher Education Institutions in terms of staff time was £110 to £165 million per year for peer review and up to £30 million per year for the work done by editors and editorial boards [4]. Worldwide, peer review costs an estimated £1.9 billion annually and accounts for about one-quarter of the overall costs of scholarly publishing and distribution [5]. The human cost was evaluated at 15 million hours by 2010 [6]. Furthermore, the peer review system may fail to identify

* Correspondence: isabelle.boutron@htd.aphp.fr
†Equal contributors
[1]Centre d'Épidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique des Hôpitaux de Paris, Paris, France
[2]Paris Descartes University, Paris, France
Full list of author information is available at the end of the article

BioMed Central

© 2015 Chauvin et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

important flaws in the quality of manuscripts and published articles [7, 8]. A recent study by Hopewell et al. [9] showed that peer reviewers could not detect important deficiencies in the reporting of methods and results of randomized trials; modifications requested by peer reviewers were relatively minor. For example, reviewers added information about concealment of allocation sequence for 10 % of the manuscripts while it was not reported in more than half of the articles. Although most had a positive impact, some were inappropriate and could have a negative impact on reporting in the final publication. Peer reviewers' assessments can also be affected by the study results: peer reviewers tend to rate the methodological quality of a study with positive results higher than the same study reported with negative results [10, 11]. Moreover, the reproducibility between reviewers is disputable [12–14]. Finally, peer reviewers are not able to detect fraud [15] and mistakes [16].

Our hypothesis is that several tasks can be expected from peer reviewers and that peer reviewers and editors may have different views on which task is important.

We aimed to 1) identify all tasks that are expected of peer reviewers when evaluating a manuscript reporting a randomized controlled trial (RCT), 2) sort these tasks according to their importance, and 3) determine which tasks are clearly requested by editors in their recommendations to reviewers. We focused on the peer review of RCTs because this design is the cornerstone of therapeutic evaluation.

## Methods

We proceeded in three steps: first, we identified and combined the tasks expected of peer reviewers, then surveyed peer reviewers of RCT reports to sort the different combined tasks, followed by assessment of the editors' recommendations to peer reviewers to determine which of the tasks identified are clearly requested.

### Identification of tasks expected from peer reviewers

To identify the tasks expected of peer reviewers, we systematically reviewed the published literature and searched for the recommendations to reviewers in a sample of journals.

#### Systematic review of the published literature

We systematically searched MEDLINE via PubMed, Google Scholar, and EMBASE for all articles related to the peer review process and reporting tasks expected of peer reviewers published in English during the last 10 years with an abstract available. The search strategy used the keywords "*peer review*" OR "*peer-review*" (Search date: March 18, 2014). One researcher (AC) read all titles and abstracts and all text fragments from full-text articles. If a study was potentially eligible, the full-text article was retrieved. We included articles in the field of biomedical research dedicated to the peer review process and specifically pertaining to tasks expected of peer reviewers.

#### Journal recommendations to reviewers

We selected a sample of journals publishing full-text articles in English. We aimed to select a large variety of journals publishing RCT results. For this purpose, we proceeded in two steps. 1) We searched PubMed with the keywords *randomized controlled trial* with the following filters: study type RCT, English language, human studies, published between January 1 and March 31, 2013, and with an abstract available (search date February 2, 2014). All citations retrieved were exported to an EndNote file. We selected all journals whatever the category (Additional file 1) that had published at least 15 reports of RCTs indexed in PubMed during the 3-month period. 2) We arbitrarily selected, in the Web of Science (Journal Citation Reports), 14 categories in the field of medicine among the 111 existing categories in the field of (bio)-medicine. We chose these categories to have a sample of journals that focused on a large variety of patients (children, adults) and treatments (drugs, surgery, rehabilitation, psychotherapy, radiotherapy, devices, etc.).The selected categories were emergency medicine, surgery, internal medicine, anesthesiology, cardiac and vascular disease, critical care medicine, hematology, infectious diseases, neurosciences, oncology, pediatrics, psychiatry, rheumatology, and clinical neurology. Then, for each category, we selected the 10 highest-impact-factor journals (Journal Citation Report 2012) that published RCTs (Additional file 1). For this purpose, we hand-searched the three last issues of each journal (2014/04/30) and selected only journals that had published at least one RCT report in the three last issues.

For each selected journal, the recommendations to peer reviewers posted on the journal website were searched (by AC) and all the editors and managing editors with an available email address were emailed to ask for the recommendations they provide to their peer reviewers. When recommendations were provided on the journal website and the journal publisher website, we combined the recommendations and considered that the peer reviewers had to follow both recommendations. One reminder was sent every week until three emails were sent.

#### Extraction and condensation of tasks

From all sources (articles, website, journal editors' responses), one researcher (AC) systematically recorded all peer review tasks relevant for the assessment of an RCT reported. Duplicates were deleted and tasks were classified into the following categories used in a previous work [17]: 1) logistics (i.e., how to use the manuscript review system), 2) etiquette/objectives (i.e., how to behave during the peer review process), 3) abstract (i.e., form and content), 4)

introduction (i.e., context of the study), 5) rationale (i.e., interest of the study), 6) methods (i.e., quality of study methods), 7) statistics (i.e., sample size calculation or proposition of a specific statistics review), 8) results (i.e., presentation and content of results), 9) discussion (i.e., interpretation of results and limitations), 10) conclusion (i.e., whether the conclusion is supported by results), 11) references (i.e., pertinence and presentation of references), 12) overall presentation (i.e., language, length and organization of the article), 13) figures/tables (i.e., presentation, size and number of figures), 14) reporting guidelines (i.e., reporting guidelines checklist such as CONSORT), 15) ethics (i.e., reference to ethics review board or respecting patient rights), and 16) fraud (i.e., risk of plagiarism or conflict of interest) (Additional file 2).

All tasks were secondarily combined to create principal task combinations [18, 19] (Fig. 1). In fact, some of the tasks collected were defined precisely and were related to the same domain. For example, evaluating the risk of bias could imply several tasks: evaluating the quality of the randomization procedure, blinding of patients, care providers, outcome assessors, and rating and handling of missing data. Similarly, assessing the quality of randomization procedure can be divided into different tasks such as evaluating the allocation sequence generation and allocation concealment, which imply checking who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions and evaluating the mechanism used to implement the random allocation sequence (such as sequentially numbered containers), assessing any steps taken to conceal the sequence until interventions were assigned. We did not want to go into such details for defining a task, so we condensed all the tasks related to

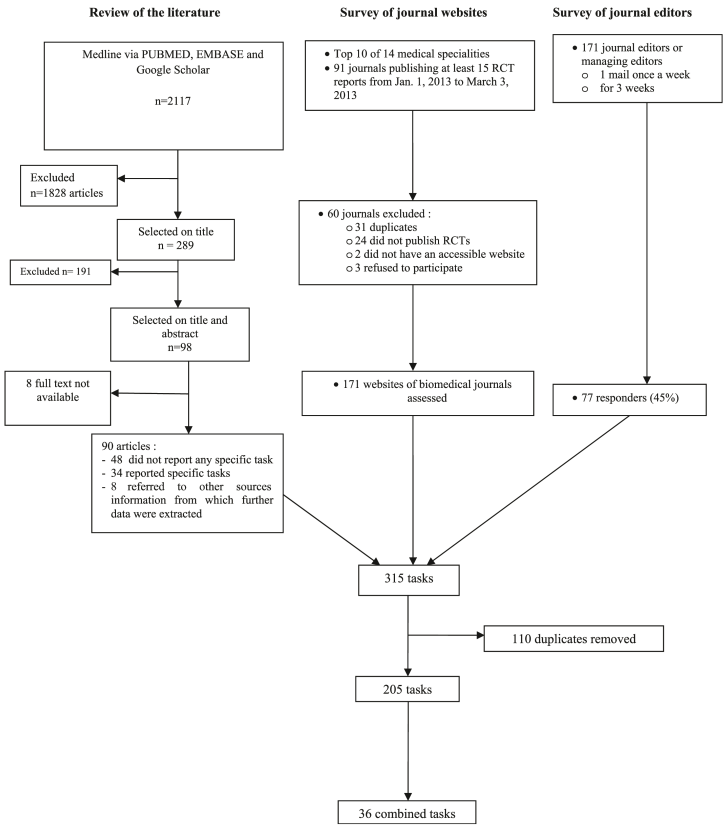



**Fig. 1** Flow chart of the survey of tasks expected of peer reviewers

the assessment of the risk of bias into a single task. We proceeded similarly for all tasks that needed some condensation.

Two researchers (AC and CB) independently combined the tasks, and consensus was reached by discussion in case of disagreement. If no consensus was reached, a third researcher helped resolve any disagreements (IB).

#### Survey of peer reviewers to sort the statements related to tasks by importance

To sort the tasks expected from peer reviewers, we used Q-sort [20, 21], which asks participants to sort combined tasks in relation to each other according to a predetermined distribution.

#### Participants

We selected participants who had peer-reviewed at least one RCT report. Our aim was to obtain the participation of peer reviewers with different expertise (methodological and clinical expertise) and backgrounds. For this purpose we used a large strategy of recruitment and contacted the following participants:

- A convenience sample of participants listed as peer reviewers of the *New England Journal of Medicine*, *Journal of the American Medical Association*, and *Annals of the Rheumatic Diseases* cited by the journal in the last journal issues for 2010 and 2011 and selected in a previous study [22].
- Corresponding authors of 1) articles published in the 10 highest-impact-factor epidemiologic journals indexed in *Public, Environmental & Occupational Health* (Journal Citation Report 2012) and in the biostatistical medical journals *Statistics in Medicine, Biostatistics, Biometrics,* and *Statistical Methods in Medical Research* and 2) RCT articles indexed in PubMed between January 1 and June 15, 2014 (date of search: June 16, 2014 search strategy). For each article selected, we extracted the corresponding author's email address, when available.
- Editors of the journals selected by searching journal websites described previously.

All participants were invited via a standardized and personalized email to participate in a survey to determine which tasks are most important when reviewing the report of an RCT. There was no reminder sent after the first email. We included only participants who reported in the survey that they had peer-reviewed at least one RCT report.

#### Sorting tasks

With Q-sort, participants classify the importance of each task relative to other tasks [23]. We pilot-tested the survey with 10 researchers (two statisticians, six methodologists, and two clinicians) and improved the wording of some items. Sorting involved three steps. First, participants read a list of tasks identified and combined in a previous step that could be expected of a peer reviewer when evaluating the report of an RCT. Second, participants sorted the combined tasks into three categories according to their evaluation of the task's importance: 1) less important, 2) neutral, or 3) more important. Third, participants were asked to read the sorted tasks and place them on a score sheet representing a normal distribution ranging from −5, less important, to +5, more important. An example of sorting is given in Additional file 3. Finally, participants could re-evaluate their distribution and shift items. We used FlashQ Software 1.0, the Flash application for performing Q-sort online [24].

#### Assessment of tasks requested in editors' recommendations to peer reviewers

For all tasks identified and sorted, we systematically determined whether the tasks were clearly recommended by journal editors. One of us (AC) screened all journal websites and documents sent by editors to determine whether editors specifically request these tasks. If the task requested by editors was not clear, we considered that editors did not request this task. For example, if editors requested "to evaluate the quality of the study," we considered this task vague or unclear. However, if editors clearly requested the assessment of the randomization procedure or the strategy used to account for missing data or a blinding procedure, we considered the editor's request clear, in this case, requesting the evaluation of risk of bias.

We also systematically recorded whether the guidelines were easily accessible (i.e., information accessible on the main page of the website), if the journal's website had a specific peer review section, and if the website included a description of the peer review process.

#### Statistical analysis

Quantitative data are described with means and SDs and categorical data with frequencies and percentages. Participant characteristics and classifications of tasks expected of peer reviewers are described overall and by participant competence (i.e., clinician, methodologist, or both). The classification was established by the mean ranking and proportion of participants who ranked the tasks in the first tertile in importance (i.e., ≥2 on the −5 to + 5 scale). We also performed a sensitivity analysis excluding the sub-group of peer reviewers who were first contacted because they were editors. These participants were first asked to provide details of their journal recommendations to reviewers and then asked again to rank the order of importance of the different tasks (Additional file 4). Statistical analysis involved use of SAS 9.3 (SAS Inst. Inc., Cary, NC).

## Results

### Identification of tasks expected from peer reviewers

We retrieved 2 117 citations from our search of articles related to tasks for peer reviewers; 98 full-text articles were selected and 90 were evaluated. Overall, among the 90 articles selected and evaluated, 48 were secondarily excluded because they did not report any recommendation or tasks dedicated to peer reviewers, 34 reported specific recommendations and tasks, and 8 referred to other sources (such as online training resources, website, etc.) from which further data were extracted (Fig. 1). The selected articles are in Additional file 5.

Overall, we selected 171 journals. All journals had a website. In total, 95 (56 %) journals did not have recommendations for peer reviewers available on their website and publisher interface; 34 (20 %) had recommendations only on their website, 32 (18 %) only on the publisher interface, and 10 (6 %) on both. Only 52 journals (30 %) had a specific peer reviewer section. Of the 171 journal editors contacted, 77 (45 %) responded and 69 (40 %) agreed to send us the documents they send to reviewers (Fig. 1), but three refused because they considered this information confidential. Among those who sent us documents, 44 (57 %) provided specific documents to their peer reviewers that are not available on their website. Among these specific documents, only 22 (50 %) gave more information than the website.

From all sources evaluated and after deletion of duplicates, we recorded 205 tasks (Fig. 1; Additional file 2). These tasks were combined into 36 principal tasks. These items concerned the following domains: etiquette/objectives (n = 3), abstract (n = 2), rationale (n = 2), methods (n = 6), statistics (n = 2), results (n = 4), discussion (n = 4), conclusions (n = 1), figures/tables (n = 2), references (n = 1), reporting guidelines (n = 2), trial registration (n = 2), presentation (n = 2), ethics (n = 1), and fraud (n = 2) (Table 1).

### Survey of peer reviewers to sort the statements related to tasks by importance

#### *Participants*

We identified 7 996 email addresses useful for inviting participants to sort tasks; 229 participants completed the Q-sort survey of which 26 were excluded because they had not peer-reviewed at least one RCT report. Therefore, 203 were included in the final analysis. The responder proportions were as follows: 21 (10.3 %) from the editors panel, 40 (19.7 %) from the reviewers panel, 142 (70 %) from corresponding authors, 57 (40 %) being corresponding authors of *Public Environmental and Occupational Health* and other biostatistical journals, and 85 corresponding authors of RCTs indexed in PubMed.

Participant characteristics are in Table 2. Participants were mainly located in Europe (49 %) and USA/Canada (32.5 %) but also in South America (9.6 %) and Oceania (8.9 %). In total, 93 (45.8 %) had clinical expertise, 72 (35.5 %) were methodologists or statisticians, 17 (8.4 %) had both clinical and methodological expertise, and 21 (10.3 %) had expertise in another area (i.e., were researchers, engineers, economists). Most participants (92 %) were investigators of at least one RCT. Overall, 150 participants (73.9 %) reviewed one to five RCT reports per year. One third did not receive any form of training to perform a peer review. More than half of the participants (61.1 %, n = 124) spent more than 2 h peer-reviewing a manuscript. Half regularly agreed to disclose their name when proposed or requested by editors.

#### Sorting tasks

Table 3 reports the tasks sorted by participants. The nine tasks rated as the most important were to 1) evaluate the risk of bias of the trial, 2) determine whether the manuscript conclusion is consistent with the results, 3) evaluate the adequacy of the statistical analyses, 4) evaluate whether the control group is appropriate, 5) check if all outcomes are adequately reported, 6) evaluate the relevance of the primary outcome(s), 7) search for any attempt to distort the presentation or interpretation of results, 8) evaluate the reliability and validity of the outcome measures, and 9) evaluate the importance of the study. For clinicians, the 2nd and 3rd most important tasks were to evaluate the importance of the study and the reliability and validity of the outcome measures, whereas for methodologists, these tasks were ranked lower, 14th and 11th, respectively. The task to evaluate the adequacy of statistical analyses was ranked 2nd by methodologists but 6th by clinicians. The sensitivity analysis excluding the answers of the 21 participants invited because they were editors showed consistent results (see Additional file 4).

### Assessment of tasks requested in editors' recommendations to peer reviewers

The tasks rated in the first tertile of importance (≥2 on the −5 to +5 scale) were not completely congruent with the tasks most frequently requested by editors (Fig. 2). For example, evaluating the risk of bias was rated in the first tertile by 63 % of participants but clearly requested by only 5 % of editors. In contrast, the task most frequently requested by editors was the recommendation for publication (76 %), but this task was classified in the first tertile by only 21 % of all participants.

Among the 171 websites studied, only 76 (44 %) had recommendations for peer reviewers. These recommendations were easily accessible for only half of the websites (n = 38). One third had had a specific peer review section. However, 70 % (124) included a description of the peer review process.

**Table 1** Thirty-six tasks created for the Q-sort survey

| | |
|---|---|
| Etiquette objectives | To read the journals' recommendations to reviewers |
| | To provide recommendations on publication (e.g., reject/revise/publish) |
| | To evaluate all appendices when available |
| Rationale | To evaluate the novelty of the study (i.e., does the trial add enough to what is already in the published literature) |
| | To evaluate the importance of the study (i.e., usefulness for clinical practice) |
| Methods | To evaluate if the control group is appropriate |
| | To evaluate the risk of bias of the trial |
| | To evaluate the adequacy of the selection of participants and clinical setting |
| | To check if the intervention is described with enough details to allow replication |
| | To evaluate the relevance of the primary outcome(s) |
| | To evaluate the reliability and validity of the outcome measures |
| Trial registration | To compare information recorded on a clinical trials register such as ClinicalTrials.gov and reported in the manuscript |
| | To compare information recorded in the trial protocol when provided by the authors and reported in the manuscript |
| Reporting guidelines | To check if the items requested by the CONSORT statement are adequately reported by authors |
| | To check if items requested by the CONSORT extensions (e.g., cluster, non-pharmacologic treatments etc.) are adequately reported when appropriate |
| Ethic | To check if the study reported ethics review board approval |
| Statistics | To evaluate the adequacy of statistical analyses |
| | To check the sample size calculation |
| Results | To search for any inconsistencies or errors in the manuscript |
| | To search for any attempt to distort the presentation or interpretation of results (e.g., data "beautification", spin, selective reporting) |
| | To check if all outcomes are adequately reported (results for each group, and the estimated effect size and its precision such as 95 % confidence interval) |
| | To check if all adverse events are adequately reported (participant withdrawals due to harms, absolute risk per arm and per adverse event type, grade, and seriousness) |
| Discussion | To evaluate if the discussion is consistent with the results |
| | To check if the authors referenced all important studies |
| | To check that limitations are adequately reported |
| | To discuss the results in relation to other studies |
| Conclusion | To determine whether the manuscript conclusion is consistent with the results |
| Fraud | To search for plagiarism or imitation in the paper |
| | To evaluate if the manuscript can be suspected of fraud |
| Figures tables | To check if all figures and tables are consistent with the text |
| | To evaluate whether figures and tables can be understood without having to refer the text |
| References | To evaluate if authors respect the requested format for references |
| Presentation | To evaluate the adequacy of the language (grammar, style, misspelling) |
| | To evaluate clarity of presentation |
| Abstract | To check if the authors reported all important outcomes and adverse events in the abstract |
| | To evaluate if the abstract conclusion is consistent with the results |

## Discussion

This study involved the identification and sorting of all tasks expected from peer reviewers when evaluating the report of an RCT. We performed a large systematic review of the literature and editors' guidelines and surveyed peer reviewers from various backgrounds who evaluated peer reviewer tasks by the Q-sort method.

Overall, we identified 221 different tasks that could be expected from peer reviewers when evaluating an RCT report. These tasks involved different levels of expertise. For example, assessing the adequacy of statistical analysis involves methodological and statistical expertise, whereas assessing the adequacy of the selection of participants and clinical setting involves content expertise. The tasks rated as most important were related to the methodology (evaluating the risk of bias of the trial, whether the control group is appropriate, and the reliability and validity of the outcome measures), statistics (evaluating the adequacy of statistical analyses), and results (determining whether the manuscript conclusions are consistent with the results and whether all outcomes are adequately reported, and searching for any attempt to distort the presentation or interpretation of results). We found some differences in how tasks were sorted according to the peer reviewers' expertise (clinicians, methodologists, or both). Therefore, peer reviewers will place different values on different tasks according to their expertise.

We found some differences between tasks clearly requested by editors in their recommendations and tasks rated as important by reviewers. Editors frequently ask whether an article is suitable for publication, but reviewers ranked this task 22nd among the 36 tasks. Editors had more recommendations about the format (tables/figures/presentation) and the importance of the study than the methodology. In contrast, the tasks considered the least important by peer reviewers mainly related to the format (evaluating whether authors respect the requested format for references and the adequacy of the language).

**Table 2** Participant characteristics

| Characteristic | Total N = 203 |
|---|---|
| Expertise | N (%) |
| – Clinician | 93 (45.8) |
| – Methodologist | 72 (35.5) |
| – Both (clinician/methodologist) | 17 (8.4) |
| – Other | 21 (10.3) |
| Affiliation[a] | |
| – Non-profit | 179 (95.7) |
| – For-profit | 5 (2.7) |
| – Publisher | 3 (1.6) |
| Country[b] | |
| – Oceania | 14 (8.9) |
| – South America/Asia | 15 (9.6) |
| – USA/Canada | 51 (32.5) |
| – Europe | 77 (49.0) |
| Total no. of completed randomized trials participated in as an investigator[c] | |
| –0 | 17 (8.4) |
| –1–5 | 99 (48.8) |
| –6–10 | 42 (20.7) |
| –11–15 | 19 (9.4) |
| –16–20 | 4 (2.0) |
| – >20 | 22 (10.8) |
| Mean no. of articles peer-reviewed per year[c] | |
| –1–5 | 35 (17.2) |
| –6–10 | 51 (25.1) |
| –11–20 | 55 (27.1) |
| –20–50 | 39 (19.2) |
| – >50 | 23 (11.3) |
| Mean no. of articles reporting an RCT peer-reviewed per year[c] | |
| –1–5 | 150 (73.9) |
| –6–10 | 27 (13.3) |
| –11–20 | 13 (6.4) |
| –20–50 | 8 (3.9) |
| – >50 | 5 (2.5) |
| Ask colleagues to help with the peer review[c] | |
| – Never | 58 (28.6) |
| – Rarely | 92 (45.3) |
| – Sometimes | 41 (20.2) |
| – Regularly | 10 (4.9) |
| – Always | 2 (1.0) |
| Training[d] | |
| – Formal academic training | 33 (16.4) |
| – Mentoring by your supervisor | 59 (29.4) |
| – Tutorial on training sessions offered by editors | 13 (6.5) |
| – Not trained | 61 (30.3) |
| – Other trained | 2 (1.0) |
| – Several forms of training | 33 (16.4) |
| Mean time for a peer review of an RCT (hours)[c] | |
| – <1 | 7 (3.4) |
| –1–2 | 72 (35.5) |
| –2–4 | 85 (41.9) |
| – >4 | 39 (19.2) |
| Agree to disclose your name when proposed or requested by an editor[c] | |
| – Never | 25 (12.3) |
| – Rarely | 25 (12.3) |
| – Sometimes | 49 (24.1) |
| – Regularly | 45 (22.2) |
| – Always | 59 (29.1) |

[a]12 (5.9 %) missing data; [b]46 (22.7 %) missing data; [c]No missing data; [d]2 (1 %) missing data
*RCT* Randomized controlled trial

Our findings also reflect that current recommendations by editors may not be sufficiently explicit. Editors' recommendations were mainly generic and not for a specific type of study design. This is probably related to the fact that editors usually rely on the expertise of reviewers and expect that the reviewers know which aspects are important and which aspects need to be evaluated. However, some evidence in the literature shows that the impact of peer reviewers is questionable [9, 24–29]. This situation is problematic because peer reviewers are usually not specifically trained; they have limited time for peer review and receive minimal rewards for this task. Further, they may misunderstand what they are expected to do. Moreover, there may be ambiguities regarding who should perform a specific task. Some tasks, such as checking the quality of reporting or checking trial registration, can be considered by editors as a reviewer task and by reviewers as an editorial task. Editors could try to achieve consensus on what they expect from peer reviewers when evaluating the report of an RCT.

Furthermore, important and simple tasks that could easily increase a paper's quality were poorly rated by peer reviewers and rarely requested by editors. In fact, the task dedicated to trial registration (i.e., the item to compare information recorded on a clinical trial register such as ClinicalTrials.gov and reported in the manuscript) was classified 32nd among the 36 condensed items and never clearly requested by editors. This situation raises some concern because trial registration may limit the risk of selective reporting bias [30]. However, these results

**Table 3** Tasks sorted by participants in the Q-sort survey. Data are (overall rank) and mean ± SD rank of participants

| | Total (N = 203) | Clinician (N = 93) | Methodologist (N = 72) | Both (N = 17) | Other (N = 21) |
|---|---|---|---|---|---|
| To evaluate the risk of bias of the trial | **(1)** 2.1 ± 2.0 | **(5)** 1.3 ± 2.1 | **(1)** 2.8 ± 1.8 | **(1)** 3.6 ± 0.8 | **(1)** 2.5 ± 1.6 |
| To determine if the manuscript conclusion is consistent with the results | **(2)** 1.9 ± 1.9 | **(1)** 1.7 ± 2.0 | **(3)** 2.1 ± 1.8 | **(2)** 2.4 ± 2.1 | **(3)** 1.9 ± 1.9 |
| To evaluate the adequacy of statistical analyses | **(3)** 1.8 ± 1.7 | **(6)** 1.2 ± 1.6 | **(2)** 2.6 ± 1.7 | **(3)** 2.2 ± 1.6 | **(5)** 1.6 ± 1.5 |
| To evaluate if the control group is appropriate | **(4)** 1.4 ± 1.8 | **(8)** 1.1 ± 1.7 | **(4)** 1.9 ± 1.8 | **(8)** 1.2 ± 1.4 | **(6)** 1.4 ± 2.1 |
| To check if all outcomes are adequately reported | **(5)** 1.4 ± 2.0 | **(9)** 1.1 ± 2.1 | **(5)** 1.7 ± 1.9 | **(5)** 1.5 ± 1.6 | **(4)** 1.9 ± 1.9 |
| To evaluate the relevance of the primary outcome(s) | **(6)** 1.3 ± 2.1 | **(4)** 1.3 ± 2.1 | **(6)** 1.5 ± 2.1 | **(14)** 0.6 ± 2.6 | **(8)** 1.1 ± 1.7 |
| To search for any attempt to distort the presentation or interpretation of results | **(7)** 1.2 ± 2.0 | **(11)** 0.9 ± 2.0 | **(7)** 1.3 ± 1.8 | **(4)** 1.9 ± 2.5 | **(2)** 2.2 ± 2.1 |
| To evaluate the reliability and validity of the outcome measures | **(8)** 1.1 ± 2.0 | **(3)** 1.4 ± 2.2 | **(11)** 0.9 ± 1.9 | **(12)** 0.7 ± 2.1 | **(9)** 1.0 ± 1.9 |
| To evaluate the importance of the study | **(9)** 1.1 ± 2.4 | **(2)** 1.7 ± 2.4 | **(14)** 0.6 ± 2.5 | **(17)** -0.1 ± 2.0 | **(13)** 0.7 ± 2.0 |
| To evaluate if the abstract conclusion is consistent with the results | **(10)** 1.0 ± 1.9 | **(10)** 1.0 ± 2.1 | **(9)** 1.2 ± 1.7 | **(9)** 1.2 ± 1.9 | **(20)** 0.0 ± 1.7 |
| To evaluate if the discussion is consistent with the results | **(11)** 1.0 ± 1.8 | **(7)** 1.2 ± 1.7 | **(13)** 0.7 ± 1.7 | **(13)** 0.7 ± 1.9 | **(7)** 1.2 ± 2.1 |
| To check if all adverse events are adequately reported | **(12)** 0.9 ± 1.9 | **(14)** 0.7 ± 2.0 | **(8)** 1.2 ± 1.7 | **(6)** 1.4 ± 1.9 | **(10)** 0.9 ± 2.1 |
| To check if the intervention is described with enough details to allow replication | **(13)** 0.8 ± 1.8 | **(16)** 0.6 ± 1.8 | **(10)** 1.1 ± 1.7 | **(10)** 1.0 ± 1.8 | **(18)** 0.2 ± 1.9 |
| To check that limitations are adequately reported | **(14)** 0.7 ± 1.7 | **(17)** 0.4 ± 1.8 | **(12)** 0.8 ± 1.6 | **(7)** 1.3 ± 1.7 | **(11)** 0.9 ± 1.6 |
| To evaluate the adequacy of the selection of participants and clinical setting | **(15)** 0.5 ± 1.9 | **(13)** 0.7 ± 1.9 | **(15)** 0.5 ± 1.8 | **(15)** 0.4 ± 1.7 | **(17)** 0.2 ± 2.0 |
| To search for any inconsistencies or errors in the manuscript | **(16)** 0.3 ± 2.2 | **(21)** 0.0 ± 2.1 | **(17)** 0.3 ± 2.2 | **(11)** 1.0 ± 2.4 | **(14)** 0.6 ± 2.0 |
| To evaluate the novelty of the study | **(17)** 0.2 ± 2.4 | **(12)** 0.8 ± 2.5 | **(20)** −0.1 ± 2.0 | **(21)** −0.3 ± 2.1 | **(24)** −0.5 ± 2.7 |
| To check the sample size calculation | **(18)** 0.2 ± 2.1 | **(18)** 0.4 ± 1.9 | **(18)** −0.1 ± 2.4 | **(24)** −0.5 ± 2.0 | **(15)** 0.5 ± 2.1 |
| To check if the authors reported all important outcomes and adverse events in the abstract | **(19)** 0.1 ± 2.1 | **(22)** 0.0 ± 2.3 | **(16)** 0.4 ± 1.9 | **(20)** −0.2 ± 1.8 | **(23)** −0.3 ± 2.3 |
| To discuss the results in relation to other studies | **(20)** 0.0 ± 1.9 | **(19)** 0.3 ± 2.0 | **(19)** −0.1 ± 1.6 | **(30)** −1.0 ± 2.1 | **(19)** 0.1 ± 2.0 |
| To evaluate if the manuscript can be suspected of fraud | **(21)** −0.1 ± 2.6 | **(15)** 0.7 ± 2.7 | **(30)** −1.1 ± 2.4 | **(29)** −0.9 ± 1.6 | **(12)** 0.9 ± 2.3 |
| To provide recommendations on publication | **(22)** −0.2 ± 2.5 | **(20)** 0.3 ± 2.8 | **(23)** −0.5 ± 2.2 | **(22)** −0.4 ± 2.3 | **(26)** −0.7 ± 2.3 |
| To check if all figures and tables are consistent with the text | **(23)** −0.2 ± 1.7 | **(23)** −0.1 ± 1.7 | **(21)** −0.4 ± 1.6 | **(19)** −0.2 ± 2.0 | **(21)** −0.1 ± 1.6 |
| To evaluate clarity of presentation | **(24)** −0.5 ± 1.9 | **(26)** −0.5 ± 2.0 | **(22)** −0.5 ± 1.9 | **(28)** −0.8 ± 1.6 | **(27)** −0.7 ± 1.7 |
| To check if the study reported ethics review board approval | **(25)** −0.5 ± 2.2 | **(25)** −0.3 ± 2.3 | **(27)** −0.9 ± 2.1 | **(23)** −0.4 ± 2.1 | **(22)** −0.3 ± 2.3 |
| To search for plagiarism or imitation in the paper | **(26)** −0.7 ± 2.3 | **(24)** −0.1 ± 2.4 | **(33)** −1.4 ± 2.2 | **(32)** −1.4 ± 2.0 | **(16)** 0.2 ± 1.9 |
| To compare information recorded in the trial protocol when provided by the authors and reported in the manuscript | **(27)** −0.7 ± 2.3 | **(27)** −0.8 ± 2.2 | **(24)** −0.6 ± 2.5 | **(18)** −0.2 ± 2.2 | **(30)** −1.0 ± 2.1 |
| To check if the items requested by the CONSORT Statement are adequately reported by authors | **(28)** −0.9 ± 2.3 | **(29)** −1.1 ± 2.2 | **(26)** −0.8 ± 2.2 | **(16)** 0.1 ± 2.6 | **(25)** −0.5 ± 2.6 |
| To check if the authors referenced all important studies | **(29)** −1.1 ± 1.7 | **(28)** −1.1 ± 1.8 | **(25)** −0.8 ± 1.7 | **(33)** −1.6 ± 1.6 | **(33)** −1.5 ± 1.2 |
| To evaluate whether figures and tables can be understood without having to refer the text | **(30)** −1.1 ± 1.9 | **(30)** −1.1 ± 1.9 | **(31)** −1.2 ± 1.8 | **(26)** −0.7 ± 2.0 | **(29)** −1.0 ± 2.2 |
| To check if items requested by the CONSORT extensions are adequately reported when appropriate | **(31)** −1.2 ± 1.9 | **(32)** −1.5 ± 2.0 | **(29)** −1.1 ± 1.8 | **(25)** −0.7 ± 1.6 | **(28)** −0.8 ± 2.0 |
| To compare information recorded on a clinical trials register such as ClinicalTrials.gov and reported in the manuscript | **(32)** −1.3 ± 2.1 | **(31)** −1.5 ± 2.1 | **(28)** −1.0 ± 2.3 | **(31)** −1.4 ± 1.7 | **(32)** −1.5 ± 2.3 |
| To read the journals' recommendations to reviewers | **(33)** −1.4 ± 2.0 | **(33)** −1.7 ± 1.9 | **(32)** −1.2 ± 1.7 | **(27)** −0.8 ± 2.5 | **(31)** −1.1 ± 2.6 |
| To evaluate all appendices when available | **(34)** −2.3 ± 1.6 | **(34)** −2.2 ± 1.8 | **(34)** −2.5 ± 1.4 | **(34)** −1.8 ± 1.8 | **(34)** −2.8 ± 1.1 |
| To evaluate the adequacy of the language | **(35)** −2.9 ± 1.7 | **(35)** −2.7 ± 1.9 | **(35)** −3.0 ± 1.5 | **(35)** −3.1 ± 1.3 | **(35)** −3.4 ± 1.4 |
| To evaluate if authors respect the requested format for references | **(36)** −4.0 ± 1.4 | **(36)** −3.8 ± 1.6 | **(36)** −4.0 ± 1.3 | **(36)** −4.4 ± 1.2 | **(36)** −4.0 ± 1.4 |

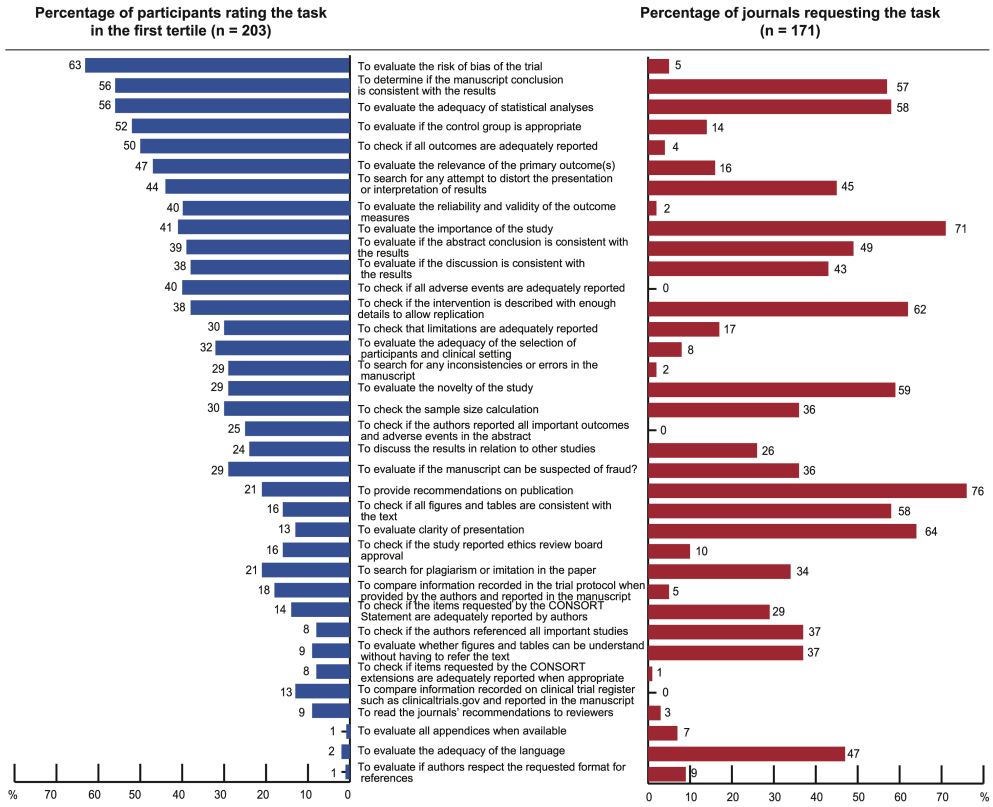



**Fig. 2** Representation for each task of the proportion of participants rating the task in the first tertile (i.e., ≥2 on the scale from −5 to +5 for each task) and the proportion of editors requesting the task in their recommendations to authors. The tasks are sorted according to the mean ranking of participants

are consistent with other studies. In fact, Mathieu et al. [22] showed that only one third of peer reviewers examined registered trial information and reported any discrepancies to journal editors. These results could be linked to the ambiguity on who, editors or reviewers, is responsible for this task. The medical community should be aware of the importance of trial registration and ensure that proper information is submitted to registries. There may be legitimate discrepancies between the manuscript and registry record, but these should be transparently reported in the paper or in the registry.

Similarly, the tasks dedicated to reporting guidelines (i.e., checking whether the items requested by the CONSORT Statement are adequately reported by authors and whether items requested by the CONSORT Statement extensions are adequately reported when appropriate) were classified 28th and 31st, respectively, among the 36 items and requested by 29 % and 1 % of editors, respectively. Some studies have demonstrated the positive impact of reporting guidelines, such as the CONSORT guidelines, on reporting quality [31]. The ranking of tasks concerning the CONSORT Statement could explain in part why reporting according to the CONSORT guidelines is poor [7, 32].

Our study has some limitations. First, we cannot know whether peer reviewers who participated in the study are representative of all peer reviewers. In fact, we cannot estimate the response rate because the emails sent may have been identified as spam or sent to junk mail folders. However, this survey focuses on qualitative input and it was mainly important to have a variety of expertise. Second, our sample included only 171 journals, with a response rate of journal editors of 45 %. Third, only one author extracted data on the tasks expected of peer reviewers (AC). However, two reviewers achieved consensus on the classification of these tasks. Fourth, our study was a

snapshot of editors' recommendations, and some journals may have since updated their websites and their instructions to peer reviewers (30 July, 2014). Finally, although we performed a systematic methodological review, we cannot exclude that we missed some reports on this topic.

## Conclusions

The tasks considered most important to peer reviewers (i.e., the methodology, statistics, and results) were frequently not clearly requested by journal editors.

## Additional files

**Additional file 1: 1.1.** Journals publishing at least 15 randomized controlled trial (RCT) reports (search date February 2, 2014) indexed in PubMed with the keywords *randomized controlled trial* with the following filters: study type RCT, English language, human studies, published between January 1 and March 31, 2013, and with an abstract available. **1.2.** The 10 journals with the highest impact factor (Journal Citation Report 2012) for 14 medical areas. Only journals publishing at least one RCT report in the three last issues were included. **1.3.** Duplicate journals.

**Additional file 2: The 205 tasks after removing duplicates.**

**Additional file 3: Example of sorting with Q-sort.**

**Additional file 4: Sensitivity analysis of task ranking with and without the editor panel.**

**Additional file 5: Articles selected during the literature review.**

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

AC participated in the protocol of the study, established the survey, and performed the literature review. GB performed the statistical analysis. IB conceived of the study and participated in its design and coordination and helped draft the manuscript. CB and PR helped draft the manuscript. All authors read and approved the final manuscript.

#### Author details

[1]Centre d'Épidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique des Hôpitaux de Paris, Paris, France. [2]Paris Descartes University, Paris, France. [3]INSERM, UMR 1153, Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité – (CRESS), METHODS team, Paris, France.

Received: 20 January 2015 Accepted: 5 June 2015
Published online: 03 July 2015

#### References

1. Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. JAMA. 1994;272:96–7.
2. Jefferson T, Rudin M, Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. Cochrane Database Syst Rev. 2007;18:MR000016.
3. Rennie R. Editorial peer review: its development and rationale. In: Godlee F, Jefferson T, editors. Peer review in health sciences. 2nd ed. London: BMJ Books; 2003. p. 1–13.
4. Public Library of Science. Peer review—optimizing practices for online scholarly communication. In: House of Commons Science and Technology Committee, editor. Peer Review in Scientific Publications, Eighth Report of Session 2010–2012, Vol. I: Report, Together with Formal, Minutes, Oral and Written Evidence. London: The Stationery Office Limited; 2011. p. 21–2.
5. Public Library of Science. Peer review—optimizing practices for online scholarly communication. In: House of Commons Science and Technology Committee, editor. Peer Review in Scientific Publications, Eighth Report of Session 2010–2012, Vol. I: Report, Together with Formal, Minutes, Oral and Written Evidence. London: The Stationery Office Limited; 2011. p. 174–8.
6. Rubriq. How we found 15 million hours of lost time. In: Rubriq: independent peer review system. 2013. http://blog.rubriq.com/2013/06/03/how-we-found-15-million-hours-of-lost-time/. Accessed 27 Mar 2015.
7. Ghimire S, Kyung E, Kang W, Kin E. Assessment of adherence to the CONSORT statement for quality of reports on randomized controlled trial abstracts from four high-impact general medical journals. Trials. 2012;13:77.
8. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA. 2010;303:2058–64.
9. Hopewell S, Collins GS, Boutron I, Yu LM, Cook J, Shanyinde M, et al. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. BMJ. 2014;349:g4145. doi:10.1136/bmj.g4145.
10. Turner EH, Matthew AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its confluence on apparent efficacy. N Engl J Med. 2008;358:252–60.
11. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine–selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ. 2003;326:1171–3.
12. Rowe BH, Strome TL, Spooner C, Bilitz S, Grafstein E, Worster A. Reviewer agreement trends from four years of electronic submissions of conference abstract. BMC Med Res Methodol. 2006;6:14.
13. Jackson JL, Srinivasan M, Rea J, Fletcher KE, Kravitz RL. The validity of peer review in a general medicine journal. PLoS ONE. 2011;6:e22475. doi:10.1371/journal.pone.0022475.
14. Weiner BK, Weiner JP, Smith HE. Spine journals: is reviewer agreement on publication recommendations greater than would be expected by chance? Spine J. 2010;10:209–11.
15. Bohannon J. Who's afraid by peer review? Science. 2013;342:60–5.
16. Schroter S, Black N, Evans S, Godlee F, Osorio L, Smith R. What errors do peer reviewers detect, and does training improve their ability to detect them? J R Soc Med. 2008;101:507–14.
17. Hirst A, Altman DG. Are peer reviewers encouraged to use reporting guidelines? A survey of 116 health research journals. PlosOne. 2012;7:e35621.
18. Van Exel J, de Graaf G. Q methodology: A sneak preview. http://qmethod.org/articles/vanExel.pdf. 2005. Accessed 28 Mar 2014.
19. Watts S, Stenner P. Doing Q methodological research: theory, method and interpretation. London: Sage Publications; 2012.
20. Brown S. Q-methodology and qualitative research. Qual Life Res. 1996;6:561–7.
21. Dennis KE. Q methodology: relevance and application to nursing research. Adv Nurs Sci. 1986;8:6–17.
22. Mathieu S, Chan AW, Ravaud P. Use of trial register information during the peer review process. PLoS One. 2013;10:8(4).
23. Stephenson W. Correlating persons instead of tests. J Personality. 1935;4:17–24. doi:10.1111/j.1467-6494.1935.tb02022.x.
24. Hackert C, Braehler G. FlashQ. 2007. http://www.hackert.biz/flashq/. Accessed 30 Mar 2014.
25. Baxt WG, Waeckerle JF, Berlin JA, Callaham ML. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. Ann Emerg Med. 1998;32:310–7.
26. Kravitz RL, Franks P, Feldman MD, Gerrity M, Byrne C, Tierney WM. Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? PLoS One. 2010;5:e10072.
27. Henderson M. Problems with peer review. BMJ. 2010;340:c1409. doi:http://dx.doi.org/10.1136/bmj.c1409.
28. Yaffe MB. Re-reviewing peer review. Sci Signal. 2009;2:eg11.
29. Stahel PF, Moore EE. Peer review for biomedical publications: we can improve the system. BMC Med. 2014;12:179.
30. Matthieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA. 2009;302:977–84.
31. Cobo E, Cortés J, Ribera JM, Cardellach F, Selva-O'Callaghan A, Kostov B, 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;22:343.
32. Mills EJ, Wu P, Gagnier J, Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. Contemp Clin Trials. 2005;26:480–7.

## 3.8 Appendices de l'article

Appendix 1.1. Journals publishing at least 15 randomized controlled trial (RCT) reports (search date February 2, 2014) indexed in PubMed with the keywords *randomized controlled trial* with the following filters: study type RCT, English language, human studies, published between January 1, 2013 and March 31, 2013, and with an abstract available

| **Journal's Name** | **Impact Factor** |
|---|---|
| Accident Analysis and Prevention | 1,964 |
| Acta Anaesthesiologica Scandinavica | 2,355 |
| Addictive Behaviors | 2,021 |
| Addiction | 4,746 |
| AIDS and Behavior | 2,979 |
| Alcoholism- Clinical and Experimental Research | 3,421 |
| American Heart Journal | 4,497 |
| American Journal of Cardiology | 3,209 |
| American Journal of Clinical Nutrition | 6,504 |
| American Journal of Kidney Diseases | 5,294 |
| American Journal of Ophtalmology | 3,631 |
| Anesthesia and Analgesia | 3,300 |
| Annals of Oncology | 7,384 |
| Antimicrobial Agents and Chemotherapy | 4,565 |
| Appetite | 2,541 |
| Archives of Gynecology and Obstetrics | 1,330 |
| Archives of Physical Medicine and Rehabilitation | 2,358 |
| Atherosclerosis | 3,991 |
| Behaviour Research and Therapy | 3,471 |
| BJOG- an International Journal Of Obstetrics and gynaecology | 3,760 |
| BJU International | 3,046 |
| Blood | 9,060 |
| BMC Musculoskeletal Disorders | 1,875 |
| BMC Psychiatry | 2,233 |
| BMC Public Health | 2,076 |
| British Journal of Anaesthesia | 4,237 |
| British Journal of Cancer | 5,082 |
| British Journal of Clinical Pharmacology | 3,578 |
| British Journal of Nutrition | 3,302 |
| Cancer | 4,901 |
| Circulation- Heart Failure | 6,684 |
| Circulation | 15,202 |

| | |
|---|---|
| Clinical Infectious Diseases | 9,374 |
| Clinical Interventions in Aging | 2,651 |
| Clinical Journal of the American Society of Nephrology | 5,068 |
| Clinical Journal of Pain | 2,552 |
| Clinical Oral Investigations | 2,200 |
| Clinical Rehabilitation | 2,191 |
| Clinical Therapeutics | 2,230 |
| Contemporary Clinical Trials | 1,597 |
| Diabetes Care | 7,735 |
| Diabetes Obesity & Metabolism | 5,181 |
| Drug and Alcohol Dependence | 3,141 |
| European Heart Journal | 14,097 |
| European Journal of Applied Physiology | 2,660 |
| European Journal of Cancer | 5,061 |
| European Journal of Clinical Nutrition | 2,756 |
| European Journal of Clinical Pharmacology | 2,741 |
| European Journal of Heart Failure | 5,247 |
| Experimental Brain Research | 2,221 |
| Journal of Affective Disorders | 3,295 |
| Journal of Alternative and Complementary Medicine | 1,464 |
| Journal of the American College of Cardiology | 14,086 |
| Journal of Applied Physiology | 2,660 |
| Journal of Clinical Endocrinology and Metabolism | 6,430 |
| Journal of Clinical Oncology | 18,038 |
| Journal of Clinical Periodontology | 3,688 |
| Journal of Clinical Pharmacology | 2,963 |
| Journal of Dentistry | 3,200 |
| Journal of Drugs in Dermatology | 1,161 |
| Journal of Infectious Diseases | 5,848 |
| Journal of Medical Internet Research | 3,768 |
| Journal of Nutrition | 4,196 |
| Journal of Pediatric | 4,035 |
| Journal of Sports Science | 2,082 |
| Journal of Strength and Conditioning Research | 1,795 |
| Journal of Substance Abuse Treatment | 1,985 |
| Journal of Urology | 3,696 |
| JAMA | 29,978 |
| Lancet | 39,060 |
| Lancet Oncology | 25,117 |
| Metabolism | 4,514 |
| New England Journal of Medicine | 51,658 |
| Neurorehabilitation and Neural Repair | 4,278 |

| | |
|---|---|
| Nicotine & Tobacco Research | 2,477 |
| Nutrition | 2,859 |
| Obesity (Silver Spring) | 3,491 |
| Osteoporosis International | 4,039 |
| Pain | 5,644 |
| Pediatrics | 5,119 |
| PLoS One | 3,730 |
| Psychology & Health | 1,950 |
| Psychopharmacology (Berl) | 4,595 |
| Respiratory Medicine | 2,585 |
| Schizophrenia Research | 4,590 |
| Stroke | 6,158 |
| Surgical Endoscopy and Other Interventional Techniques | 3,427 |
| Trials | 2,206 |
| Vaccine | 3,492 |
| Zh Nevrol Psikhiatr Im S S Korsakova | NA |
| Zhongguo Zhen Jiu | NA |

Appendix 1.2. The 10 journals with the highest impact factor (Journal Citation Report 2012) for 14 medical areas. Only journals publishing at least one RCT report in the 3 last issues were included.

| Journal's name | Impact Factor | Specialties | Selected if at least 3 RCTs of clinical trials in 3 last issues |
|---|---|---|---|
| Anaesthesia | 3,486 | Anesthesiology | Yes |
| Anesthesia and Analgesia | 3,300 | Anesthesiology | Yes |
| Anesthesiology | 5,163 | Anesthesiology | Yes |
| British Journal of Anaesthesia | 4,237 | Anesthesiology | Yes |
| European Journal of Anaesthesiology | 2,792 | Anesthesiology | Yes |
| European Journal of Pain | 3,067 | Anesthesiology | Yes |
| Minerva Anestesiologica | 2,818 | Anesthesiology | Yes |
| Pain | 5,644 | Anesthesiology | Yes |
| Pain Practice | 2,605 | Anesthesiology | Yes |
| Regional Anesthesia and Pain Medicine | 3,464 | Anesthesiology | Yes |
| Circulation | 15,202 | Cardiac & Cardiovascular systems | Yes |
| Circulation- Cardiovascular Genetics | 6,728 | Cardiac & Cardiovascular systems | Yes |
| Circulation- Cardiovascular Interventions | 6,543 | Cardiac & Cardiovascular systems | Yes |
| Circulation- Heart Failure | 6,684 | Cardiac & Cardiovascular systems | Yes |
| Circulation Research | 11,861 | Cardiac & Cardiovascular systems | Yes |
| European Heart Journal | 14,097 | Cardiac & Cardiovascular systems | Yes |
| JACC- Cardiovascular Imaging | 6,164 | Cardiac & Cardiovascular systems | Yes |
| JACC- Cardiovascular Interventions | 6,543 | Cardiac & Cardiovascular systems | Yes |
| Journal of the American College of Cardiology | 14,086 | Cardiac & Cardiovascular systems | Yes |
| Nature Reviews in Cardiology | 10,400 | Cardiac & Cardiovascular systems | No |
| Acta Neuropathologica | 9,734 | Clinical neurology | No |
| Alzheimers & Dementia | 14,483 | Clinical neurology | Yes |
| Annals of Neurology | 11,193 | Clinical neurology | Yes |
| Archives of Neurology | 7,685 | Clinical neurology | No |
| Brain | 9,915 | Clinical neurology | Yes |
| Lancet Neurology | 23,917 | Clinical neurology | Yes |
| Nature Reviews Neurology | 15,518 | Clinical neurology | No |
| Neurology | 8,249 | Clinical neurology | Yes |

| | | | |
|---|---|---|---|
| Neuro-oncology | 6,180 | Clinical neurology | Yes |
| Sleep Medicine Reviews | 8,681 | Clinical neurology | Yes |
| American Journal of Respiratory and Critical Care Medicine | 11,041 | Critical Care Medecine | Yes |
| Chest | 5,854 | Critical Care Medecine | Yes |
| Critical Care | 4,718 | Critical Care Medecine | Yes |
| Critical Care Medicine | 6,124 | Critical Care Medecine | Yes |
| Current Opinion in Critical Care | 2,967 | Critical Care Medecine | No |
| Intensive Care Medicine | 5,258 | Critical Care Medecine | Yes |
| Journal of Neurotrauma | 4,295 | Critical Care Medecine | Yes |
| Minerva Anestesiologica | 2,818 | Critical Care Medecine | Yes |
| Neurocritical Care | 3,038 | Critical Care Medecine | Yes |
| Resucitation | 4,104 | Critical Care Medecine | Yes |
| Academic Emergency Medicine | 1,757 | Emergency Medicine | Yes |
| American journal of Emergency Medecine | 1,704 | Emergency Medicine | Yes |
| Annals of Emergency Medicine | 4,285 | Emergency Medicine | Yes |
| Emergencias | 2,578 | Emergency Medicine | Yes |
| Emergency Medicine Journal | 1,645 | Emergency Medicine | Yes |
| Injury | 2,174 | Emergency Medicine | Yes |
| Journal of Trauma - Injury infection and Critical Care | 2,348 | Emergency Medicine | No |
| Prehopospital Emergency Care | 1,859 | Emergency Medicine | Yes |
| Resuscitation | 4,104 | Emergency Medicine | Yes |
| Scandinavian Journal of Trauma Resuscitation & Medicine | 1,680 | Emergency Medicine | Yes |
| Arteriosclerosis Thrombosis and Vascular Biology | 6,338 | Hematology | Yes |
| Blood | 9,060 | Hematology | Yes |
| Blood Reviews | 6,000 | Hematology | Yes |
| Circulation Research | 11,861 | Hematology | Yes |
| Haematologica - The Hematology Journal | 5,935 | Hematology | No |
| Journal of Cerebral Blood Flow and Metabolism | 5,398 | Hematology | No |
| Journal of Thrombosis and Haemostasis | 6,081 | Hematology | No |
| Leukemia | 10,164 | Hematology | Yes |
| Stem Cells | 7,701 | Hematology | No |
| Thrombosis and Haemostasis | 6,094 | Hematology | Yes |
| AIDS | 6,407 | Infectious Diseases | Yes |
| Clinical Infectious Diseases | 9,374 | Infectious Diseases | Yes |
| Current Opinion in HIV and AIDS | 4,704 | Infectious Diseases | No |
| Current Opinion in Infectious Diseases | 4,870 | Infectious Diseases | No |
| Emerging Infectious Diseases | 5,993 | Infectious Diseases | Yes |

| | | | |
|---|---|---|---|
| Eurosurveillance | 5,491 | Infectious Diseases | Yes |
| JAIDS- Journal of Acquired immune Deficinecy Syndromes | 4,653 | Infectious Diseases | Yes |
| Journal of Antimicrobial Chemotherapy | 5,338 | Infectious Diseases | Yes |
| Journal of Infectious Diseases | 5,848 | Infectious Diseases | Yes |
| Lancet Infectious Diseases | 19,966 | Infectious Diseases | Yes |
| Annals of Internal Medicine | 13,976 | Medicine, General & Internal | Yes |
| Archives of Internal Medicine | 10,579 | Medicine, General & Internal | Yes |
| BMC Medicine | 6,679 | Medicine, General & Internal | Yes |
| British Medical Journal | 17,215 | Medicine, General & Internal | Yes |
| Canadian Medical Association Journal | 6,465 | Medicine, General & Internal | Yes |
| JAMA | 29,978 | Medicine, General & Internal | Yes |
| Journal of Internal Medicine | 6,455 | Medicine, General & Internal | Yes |
| Lancet | 39,060 | Medicine, General & Internal | Yes |
| New England Journal of Medicine | 51,658 | Medicine, General & Internal | Yes |
| Plos Medicine | 15,253 | Medicine, General & Internal | Yes |
| Annals of Neurology | 11,193 | Neurosciences | Yes |
| Annual Review of Neuroscience | 20,614 | Neurosciences | No |
| Behavioral and Brain Sciences | 18,571 | Neurosciences | Yes |
| Brain | 9,915 | Neurosciences | Yes |
| Molecular Psychiatry | 14,897 | Neurosciences | Yes |
| Nature Neurosciences Neurosci | 15,251 | Neurosciences | No |
| Nature Reviews Neuroscience | 31,673 | Neurosciences | No |
| Neuron | 15,766 | Neurosciences | Yes |
| Trends in Cognitive sciences | 16,008 | Neurosciences | No |
| Trends Neurosci | 13,582 | Neurosciences | Yes |
| Biochimica et Biopsysica - Reviews of Cancer | 9,033 | Oncology | No |
| CA- A Cancer Journal for Clinicians | 153,459 | Oncology | Yes |
| Cancer Cell | 24,755 | Oncology | No |
| Cancer Discovery | 10,143 | Oncology | Yes |
| JNCI- Journal of the National Cancer Institute | 14,336 | Oncology | Yes |
| Journal of Clinical Oncology | 18,038 | Oncology | Yes |
| Lancet Oncology | 25,117 | Oncology | Yes |

| | | | |
|---|---|---|---|
| Leukemia | 10,174 | Oncology | Yes |
| Nature Reviews Cancer | 35,000 | Oncology | No |
| Nature Reviews Clinical Oncology | 14,771 | Oncology | No |
| Archives of Disease in Childhood | 3,051 | Pediatrics | Yes |
| Archives of Disease in Childhood- Fetal and Neonatal Edition | 3,451 | Pediatrics | Yes |
| Archives Pediatrics Adolescent Medicine | 4,282 | Pediatrics | Yes |
| European Child & Adolescent Psychiatry | 3,699 | Pediatrics | Yes |
| Journal of Pediatric | 4,035 | Pediatrics | Yes |
| Journal of the American Academy of Child and Adolescent Psychiatry | 6,970 | Pediatrics | Yes |
| Pediatric Allergy and Immunology | 3,376 | Pediatrics | Yes |
| Pediatrics | 5,119 | Pediatrics | Yes |
| Seminars in Fetal & Neonatal Medicine | 3,505 | Pediatrics | No |
| The Pediatric Infectious Disease Journal | 3,569 | Pediatrics | Yes |
| American journal of Psychaitry | 14,721 | Psychiatry | Yes |
| Archives of General Psychaitry | 13,772 | Psychiatry | No |
| Biological psychiatry | 9,247 | Psychiatry | No |
| British Journal of Psychiatry | 6,606 | Psychiatry | No |
| Journal of Psychiatry & Neuroscience | 6,242 | Psychiatry | No |
| Molecular Psychiatry | 14,897 | Psychiatry | Yes |
| Neuropsychopharmacology | 8,678 | Psychiatry | No |
| Psychoterapy and Psychosomatics | 7,230 | Psychiatry | No |
| Schizophrenia Bulletin | 8,486 | Psychiatry | No |
| World Psychiatry | 8,974 | Psychiatry | No |
| Annals of Rheumatic Diseases | 9,111 | Rheumatology | Yes |
| Arthritis and Rheumatology | 7,477 | Rheumatology | Yes |
| Arthritis Care & Research | 3,731 | Rheumatology | Yes |
| Arthritis Research & Therapy | 4,302 | Rheumatology | Yes |
| Best Practice & Research in Clinical Rheumatology | 3,550 | Rheumatology | Yes |
| Curr Opin Rheumatol | 5,191 | Rheumatology | No |
| Nature Reviews Rheumatology | 9,745 | Rheumatology | No |
| Osteoarthritis and Cartilage | 4,262 | Rheumatology | Yes |
| Rheumatology | 4,212 | Rheumatology | Yes |
| Seminars in Arthritis and Rheumatism | 3,806 | Rheumatology | Yes |
| American Journal of surgical pathology | 4,868 | Surgery | Yes |
| American Journal of Transplantation | 6,192 | Surgery | Yes |
| Annals of Surgery | 6,329 | Surgery | Yes |
| Annals Surgical Oncology | 4,120 | Surgery | Yes |
| Archives of Surgery | | Surgery | Yes |
| British Journal of Surgery | 4,839 | Surgery | Yes |
| Endoscopy | 5,735 | Surgery | Yes |

| | | | |
|---|---|---|---|
| Journal of neurology Neurosurgery and Psychiatry | 4,924 | Surgery | Yes |
| Journal of the American College of Surgeons | 4,500 | Surgery | Yes |
| Surgery for obesity and Related Diseases | 4,121 | Surgery | Yes |

Appendix 1.3. Duplicate journals

| |
|---|
| Annals of Neurology |
| Brain |
| Leukemia |
| Circulation Research |
| Minerva Anestesiol |
| Anesthesia and Analgesia |
| Atherosclerosis |
| Blood |
| British Journal of Anaesthesia |
| Circulation- Heart Failure |
| Circulation |
| Clinical Infectious Diseases |
| European Heart Journal |
| Journal of the American College of Cardiology |
| Journal of Clinical Oncology |
| Journal of Infectious Diseases |
| Journal of Pediatric |
| JAMA |
| Lancet |
| Lancet Oncology |
| New England Journal of Medicine |
| Pain |
| Pediatrics |
| Molecular Psychiatry |

Appendix 2. The 205 tasks after removing duplicates

| | | |
|---|---|---|
| **Logistics** | 1 | Complete the review by the assigned deadline |
| | 2 | Organize notations by sections |
| | 3 | When making critical notes on particular sections of a paper: indicate location of the section in question as specifically as possible |
| | 4 | Provide separate comments to the editors and to the author |
| **Etiquette/Objectives** | 5 | Provide constructive and useful comments |
| | 6 | Supply evidence supporting comments |
| | 7 | Provide confidential comments to editors |
| | 8 | Treat author's manuscript with respect, fairness and impartiality |
| | 9 | Be familiar with the instructions to authors, types of papers, style and standard of the journal you are reviewing for |
| | 10 | Declare your conflict of interest |
| | 11 | Maintain confidentiality concerning the existence and contents of unpublished manuscripts |
| | 12 | Do not review if reviewer has a conflict of interest |
| | 13 | Compile a neat list of recommendations for authors and editors |
| | 14 | Do not plagiarize |
| | 15 | Accept manuscripts for reviews only in yours areas of expertise |
| | 16 | Make recommendations of acceptance, rejection or revision |
| | 17 | Identify problems or limitations and help rectify them |
| | 18 | Reduce misinformation and confusion |
| | 19 | Improve the quality of manuscript |
| | 20 | Provide a list of important items that are missing in the manuscript |
| | 21 | Point out any contradictions within a manuscript |
| | 22 | Provide specific comments on the strengths and weaknesses of a submission |
| | 23 | Have expertise relevant to the research methodology |
| | 24 | Raise arguments about data in a previous publication that were already made through peer review |
| | 25 | Evaluate the theoretical framework |
| | 26 | Evaluate the scientific rigor |
| | 27 | Suggest alternative titles |
| | 28 | Evaluate if the article is relevant for the journal |
| | 29 | Protect the integrity of his or her speciality or subspeciality, the reputation of the journal, the welfare of human and animals subjects |
| | 30 | Assure that the general review quality is constructive knowledge, logical, analytical, clear/transparent, precise, balanced/fair, courteous |
| **Rational** | 31 | Review consideration of the context of previous relevant research, and the appropriateness or need for conducting the particular study |
| | 32 | Evaluate the quality of literature review and medical context |
| | 33 | Discuss the importance and originality of study |
| | 34 | Evaluate the justification for conducting the study |

| | | |
|---|---|---|
| | 35 | Evaluate the clarity of the rationale |
| | 36 | Evaluate if the logic of the rationale is clear |
| | 37 | Evaluate if hypotheses are clearly stated |
| | 38 | Evaluate if the research question is congruent with the hypotheses and appropriately described |
| | 39 | Evaluate if objectives are clearly described |
| | 40 | Check that the study is based on evidence based practice of known effectiveness |
| | 41 | Evaluate the usefulness of the research for patients and clinical practice |
| | 42 | Check if preliminary studies exist |
| | 43 | Check if competing studies exist |
| | 44 | Check if the study provides original data not available in the literature |
| | 45 | Evaluate the relevance of the study: duration/relevance |
| **Methods** | 46 | Evaluate the type of article: if it complete articulates a theory or conceptual framework |
| | 47 | Evaluate if there is congruence between the theoretical or conceptual framework and the problem |
| | 48 | Evaluate the reliability and validity of the outcome measures |
| | 49 | Evaluate the methodology quality of the study design/Scientific quality and rigor of the method |
| | 50 | Evaluate if the methods used are valid and reliable |
| | 51 | Evaluate if the design is congruent and appropriately described |
| | 52 | Evaluate if the study design is clear |
| | 53 | Evaluate the adequacy of the design and methods for the objective |
| | 54 | Evaluate if there are no serious methodological errors |
| | 55 | Identify strengths and weaknesses of methods |
| | 56 | Evaluate if authors used all techniques to avoid risk of bias |
| | 57 | Evaluate if the experimental design includes sufficient controls |
| | 58 | Evaluate if the methods are sufficiently and clearly detailed so that a reader could hypothetically reproduce the study |
| | 59 | Evaluate if the author provided rationale for the methodological choices |
| | 60 | Evaluate if the method maximizes internal and external validity as appropriate for the chosen study design |
| | 61 | Evaluate if there is no ambiguity or omission regarding materials, procedures, durations, sample sizes and selection method |
| | 62 | Evaluate if the operational definitions of the variables is clear |
| | 63 | Evaluate if the literature review is correctly performed |
| | 64 | Comment on the methods of the research |
| | 65 | Evaluate if authors use a robust method for disseminating results |
| | 66 | Evaluate if there is an adequate description of the setting and subjects |
| | 67 | Evaluate if authors use an appropriate subject population |
| | 68 | Evaluate if inclusion and exclusion criteria are clear |
| | 69 | Evaluate the relevance of inclusion criteria |

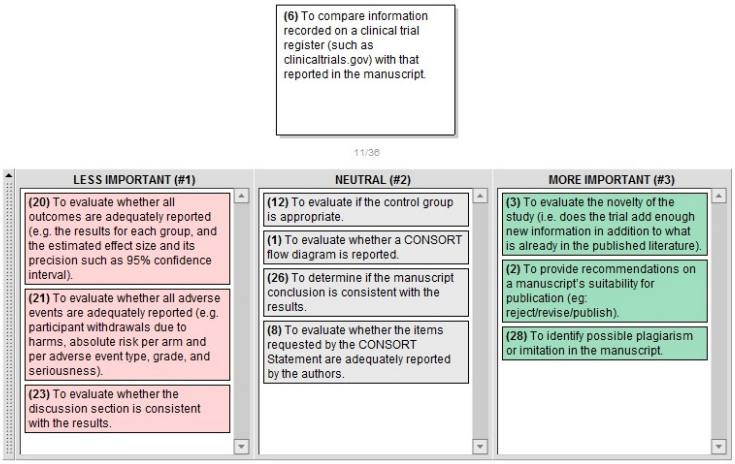
| | | |
|---|---|---|
| | 70 | Evaluate the relevance of exclusion criteria |
| | 71 | Evaluate if authors defined stopping rules |
| | 72 | Evaluate if the sample size is adequate to detect clinically meaningful effects in pre-specified primary outcomes |
| | 73 | Evaluate if the sample size calculation is reported |
| | 74 | Evaluate the randomization procedure |
| | 75 | Evaluate the relevance of randomization methods |
| | 76 | Verify that authors register his or her study in a clinical trial registry |
| | 77 | Verify that authors respect their study protocol |
| | 78 | Check If the study deviates from the protocol, verify authors explanations |
| | 79 | In case of a control group, evaluate if it is appropriate |
| | 80 | In case of a control group, evaluate if it is clearly described |
| | 81 | If authors use a new scale, evaluate if the measurement instruments are completely described |
| | 82 | If authors use a new scale, evaluate if authors explained in which population the measurement instrument was developed or validated |
| | 83 | If authors use a new scale, evaluate if authors describe factor analyses used for developing new scales and subscales |
| | 84 | If authors use a new scale, evaluate construct measurements |
| | 85 | Evaluate the quality of the description of the management of adverse events |
| | 86 | Evaluate if subgroups are relevant |
| | 87 | In case of multiple groups, evaluate if they are statistically comparable |
| | 88 | In case of a qualitative study, evaluate if the method used to collect data is appropriate to identify qualitative focus and clearly described |
| | 89 | In case of a qualitative study, evaluate if the author identified the phenomenon of interest |
| | 90 | In case of a qualitative study, evaluate if the author stated why a qualitative approach was use |
| | 91 | In case of a qualitative study, evaluate if authors discussed data saturation |
| | 92 | In case of an intervention, evaluate if it is adequately described |
| **Statistics** | 93 | Evaluate the relevance of statistical analyses |
| | 94 | Evaluate if statistical methods are appropriate |
| | 95 | Evaluate if authors justify the choice of statistic method and statistical tests |
| | 96 | Verify that authors correctly describe the population |
| | 97 | Evaluate the choice of T0 (initial time measure), duration, frequency, interval used by authors |
| | 98 | Evaluate which variables are used by authors |
| | 99 | Evaluate the statistical model, estimators and assumptions |
| | 100 | Evaluate the parameterization of statistical models |
| | 101 | Evaluate data presentation |
| | 102 | Verify that authors provide confidence intervals/ p-values/overall fit |
| | 103 | Verify if the statistical interpretation is appropriate |

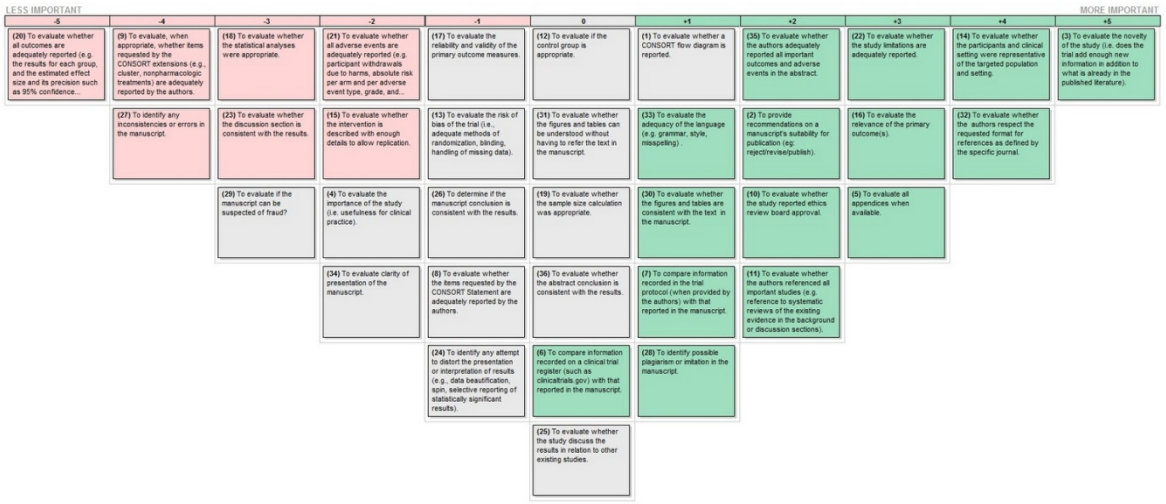
| | | |
|---|---|---|
| | 104 | Evaluate data quality |
| | 105 | Evaluate if there are statistical errors |
| | 106 | Check if authors respect intention to treat for data analyses |
| | 107 | Evaluate if authors calculated statistical power a priori |
| | 108 | Evaluate if authors used a missing data strategy |
| | 109 | Evaluate if sensitivity analyses are conducted to evaluate the impact of the missing data assumption on the results |
| | 110 | Check that the people missing from the analyses are carefully described |
| | 111 | Check if authors make multiple statistical tests, describe a rational choice regarding adjustment for alpha risk |
| | 112 | Verify that the primary analysis is in intention to treat |
| | 113 | Evaluate if authors report an interaction term between the effect and subgroup |
| | 114 | Evaluate the statistical relevance of subgroup analyses |
| | 115 | Evaluate if authors describe patient follow up |
| **Figures/Tables** | 116 | Evaluate if graphs and tables are understandable without having to refer to the text |
| | 117 | Evaluate if tables and figures summarize the data |
| | 118 | Evaluate if figures/tables make the data more understandable |
| | 119 | Evaluate if figures and tables illustrate the important features of methods and results |
| | 120 | Evaluate if figures/tables are justified |
| | 121 | Evaluate if the size of figures/tables are adequate |
| | 122 | Evaluate if the number of tables is appropriate |
| | 123 | Check if legends correctly explain the figures |
| | 124 | Evaluate if figures and tables can be simplified or condensed |
| **Results** | 125 | Evaluate if presentation of results is clear, concise and organized |
| | 126 | Check if results included in analyses includes all variables presented in the text |
| | 127 | Check if results present outcomes for each objective |
| | 128 | Check for the accuracy of results |
| | 129 | Check if authors differentiate between differences that are statistically significant and those of clinical interest |
| | 130 | Evaluate if results are reasonable and expected |
| | 131 | Evaluate if results support the conclusions drawn |
| | 132 | Evaluate if authors clearly et systematically announce the study findings |
| | 133 | Evaluate if results clearly describe the existence of missing data |
| | 134 | Evaluate if authors correctly interpret statistical analyses |
| | 135 | Evaluate if results found are not affected by confounding factors |
| | 136 | In case of therapeutic study, evaluate if authors report all adverse events |
| | 137 | In case of therapeutic study, evaluate if authors use analytic tests for adverse events |
| **Discussion** | 138 | Evaluate if authors discuss the interpretation of results |
| | 139 | Evaluate if findings are important (according to who (the authors or the reviewer?)?) |

| | | |
|---|---|---|
| | 140 | Evaluate if authors report a possible Hawthorne effect |
| | 141 | Evaluate if authors include only individuals who would meet criteria for authorship |
| | 142 | Evaluate if authors give explanations for ineligible or non-randomized cases |
| | 143 | Evaluate if study findings are clearly articulated |
| | 144 | Check that authors state whether their hypotheses were proven or not |
| | 145 | Evaluate if authors report and comment on all sources of bias and limitations |
| | 146 | Evaluate if authors compare findings and contrast findings with previous work in this area |
| | 147 | Evaluate if authors adequately synthesize the observations with those in the literature |
| | 148 | Evaluate if authors explain any unexpected results |
| | 149 | Evaluate if authors present plausible alternative explanations for their findings |
| | 150 | Evaluate if authors report all essential observations arising from the experimental design |
| | 151 | Evaluate if authors fully disclose sources of financial support |
| | 152 | Evaluate if the authors report if the findings are linked to the theoretical or conceptual framework for the study |
| | 153 | Evaluate if authors discuss how the study findings will improve patient care |
| | 154 | Evaluate if authors reassess (reassess what about the research?) the research |
| | 155 | Evaluate if the discussion is concise |
| | 156 | Evaluate if authors provide level of evidence for results |
| | 157 | Evaluate if authors overlook critical references |
| | 158 | Evaluate if the presentation of the findings misleadingly convince the reader that the data was rigorously obtained and reported |
| | 159 | Verify that there is a clear separation between exploratory (post-hoc) hypothesis-testing and primary pre planned hypothesis testing |
| | 160 | Check if authors only analyze subgroups described in the methods |
| | 161 | Assess that authors comment on the results of the research |
| **Conclusions** | 162 | Evaluate if conclusions are congruent |
| | 163 | Evaluate if conclusions are justified by the results found in study |
| | 164 | Evaluate if conclusions are appropriately described |
| | 165 | Evaluate if conclusions present new findings |
| | 166 | Evaluate if conclusions present clearly outlined major points for future research and/or clinical practice |
| | 167 | Evaluate if authors explore and explicitly state implications of the results |
| | 168 | Evaluate if authors present appropriate recommendations for future studies |
| | 169 | Evaluate if authors evaluate if research results are generalizable |
| | 170 | Evaluate if authors disproportionately interpret results |
| | 171 | Evaluate if study data permits the authors to claim efficacy |
| **Abstract** | 172 | Evaluate if the abstract reflects the manuscript as a whole, in particular the findings |
| | 173 | Evaluate if the abstract is complementary to the title |
| | 174 | Calculate the number of words of the manuscript |

| | | |
|---|---|---|
| | 175 | Check that abstract includes keywords |
| | 176 | Evaluate the relevance of keywords |
| **Introduction** | 177 | Evaluate the quality of the introduction (length, grammar, style) |
| | 178 | Check if the introduction presents epidemiological data and definitions |
| | 179 | Check if authors used definition of main terms |
| | 180 | Evaluate if the introduction argues the rationale of the current study |
| | 181 | Check if the introduction presents a literature review on the scientific and medical context |
| | 182 | Evaluate if the study is based on evidence based practice of known effectiveness |
| **References** | 183 | Evaluate if references are correctly formatted |
| | 184 | Evaluate if there are sufficient references for the purpose of this manuscript |
| | 185 | Evaluate if the references included are up to date |
| | 186 | Evaluate if there are no errors in citations |
| | 187 | Check if authors correctly use references' results |
| **Reporting guidelines** | 188 | Evaluate if authors used reporting guidelines |
| | 189 | Evaluate if the reporting guidelines are appropriate for the study |
| | 190 | Evaluate the quality of reporting guidelines |
| | 191 | Evaluate if there are errors in data reporting |
| **Ethical review** | 192 | Evaluate if the study received ethics review board approval |
| | 193 | If study did not receive ethics review board approval, evaluate the potential risks and benefits for participants, and the adequacy of the information to be provided to participants |
| | 194 | If the study did not receive ethics review board approval, evaluate the adequacy of the information provided to participants |
| | 195 | If the study did not receive ethics review board approval, evaluate the description of patient consent forms |
| | 196 | If the study did not receive ethics review board approval, evaluate the if there is a discussion regarding the rights of the study's participants |
| | 197 | If the study did not receive ethics review board approval, evaluate the research violations about humans or animals |
| | 198 | If study did not receive ethics review board approval, evaluate if this study respects the Helsinki' Declaration |
| **Fraude** | 199 | Assess the possibility of fraud |
| | 200 | Investigate for possible plagiarism or duplicate submission |
| | 201 | Report suspected authorship to editors |
| | 202 | Evaluate if authors have any conflict of Interest |
| | 203 | Check if results have been published previously |
| | 204 | Check if the manuscript has been previously published |
| | 205 | Evaluate the description of resource allocation |

Appendix 3. Example of sorting with Qsort



Appendix 4. Sensitivity analysis of tasks ranked with and without the editor panel.

| **Tasks** | **Tasks ranking with editors panel N=203** | **Tasks ranking without editors panel N=182** |
|---|---|---|
| To evaluate the risk of bias of the trial | 1 | 1 |
| To determine if the manuscript conclusion is consistent with the results | 2 | 2 |
| To evaluate the adequacy of statistical analyses | 3 | 3 |
| To evaluate if the control group is appropriate | 4 | 5 |
| To check if all outcomes are adequately reported | 5 | 4 |
| To evaluate the relevance of the primary outcome(s) | 6 | 6 |
| To search for any attempt to distort the presentation or interpretation of results | 7 | 7 |
| To evaluate the reliability and validity of the outcome measures | 8 | 8 |
| To evaluate the importance of the study | 9 | 12 |
| To evaluate if the abstract conclusion is consistent with the results | 10 | 10 |
| To evaluate if the discussion is consistent with the results | 11 | 9 |
| To check if all adverse events are adequately reported | 12 | 11 |
| To check if the intervention is described with enough details to allow replication | 13 | 13 |
| To check that limitations are adequately reported | 14 | 14 |
| To evaluate the adequacy of the selection of participants and clinical setting | 15 | 15 |
| To search for any inconsistencies or errors in the manuscript | 16 | 16 |
| To evaluate the novelty of the study | 17 | 20 |
| To check the sample size calculation | 18 | 19 |
| To check if the authors reported all important outcomes and adverse events in the abstract | 19 | 17 |
| To discuss the results in relation to other studies | 20 | 18 |
| To evaluate if the manuscript can be suspected of fraud | 21 | 21 |
| To provide recommendations on publication | 22 | 22 |
| To check if all figures and tables are consistent with the text | 23 | 23 |
| To evaluate clarity of presentation | 24 | 24 |
| To check if the study reported ethics review board approval | 25 | 25 |
| To search for plagiarism or imitation in the paper | 26 | 26 |
| To compare information recorded in the trial protocol when provided by the authors and reported in the manuscript | 27 | 27 |
| To check if the items requested by the CONSORT Statement are adequately reported by authors | 28 | 28 |
| To check if the authors referenced all important studies | 29 | 30 |
| To evaluate whether figures and tables can be understand without having to refer the text | 30 | 29 |
| To check if items requested by the CONSORT extensions are | 31 | 31 |

| | | |
|---|---|---|
| adequately reported when appropriate | | |
| To compare information recorded on clinical trial register such as clinicaltrials.gov and reported in the manuscript | 32 | 32 |
| To read the journals' recommendations to reviewers | 33 | 33 |
| To evaluate all appendices when available | 34 | 34 |
| To evaluate the adequacy of the language | 35 | 35 |
| To evaluate if authors respect the requested format for references | 36 | 36 |

Appendix 5. Articles selected during the literature review


1 Abdoul, H., et al., Peer review of grant applications: criteria used and qualitative study of reviewer practices. PLoS One, 2012. 7(9): p. e46054.

2 Alam, M., et al., Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: a randomized multi-rater study. Br J Dermatol, 2011. 165(3): p. 563-7.

3 Albers, C.A., et al., Publication criteria and recommended areas of improvement within school psychology journals as reported by editors, journal board members, and manuscript authors. J Sch Psychol, 2011. 49(6): p. 669-89.

4 Ammenwerth, E., et al., Developing and evaluating criteria to help reviewers of biomedical informatics manuscripts. J Am Med Inform Assoc, 2003. 10(5): p. 512-4.

5 Anraku, A., et al., Survey of conflict-of-interest disclosure policies of ophthalmology journals. Ophthalmology, 2009. 116(6): p. 1093-6.

6 Azer, S.A., S. Ramani, and R. Peterson, Becoming a peer reviewer to medical education journals. Med Teach, 2012. 34(9): p. 698-704.

7 Baggs, J.G., et al., Blinding in peer review: the preferences of reviewers for nursing journals. J Adv Nurs, 2008. 64(2): p. 131-8.

8 Bayne, S.C., G.P. McGivney, and S.C. Mazer, Scientific composition and review of manuscripts for publication in peer-reviewed dental journals. J Prosthet Dent, 2003. 89(2): p. 201-18.

9 Benos, D.J., et al., The ups and downs of peer review. Adv Physiol Educ, 2007. 31(2): p. 145-52.

10 Benos, D.J., K.L. Kirk, and J.E. Hall, How to review a paper. Adv Physiol Educ, 2003. 27(1-4): p. 47-52.

11 Borgwardt, S., et al., Why are psychiatric imaging methods clinically unreliable? Conclusions and practical guidelines for authors, editors and reviewers. Behav Brain Funct, 2012. 8: p. 46.

12 Bornmann, L., R. Mutz, and H.D. Daniel, A reliability-generalization study of journal peer reviews: a multilevel meta-analysis of inter-rater reliability and its determinants. PLoS One, 2010. 5(12): p. e14331.

13 Brand, R.A., Reviewing for clinical orthopaedics and related research. Clin Orthop Relat Res, 2012. 470(9): p. 2622-5.

14 Broome, M., et al., Ethical concerns of nursing reviewers: an international survey. Nurs Ethics, 2010. 17(6): p. 741-8.
15 Bydder, S., et al., Assessment of abstracts submitted to the annual scientific meeting of the Royal Australian and New Zealand College of Radiologists. Australas Radiol, 2006. 50(4): p. 355-9.
16 Callaham, M.L. and J. Tercier, The relationship of previous training and experience of journal peer reviewers to subsequent review quality. PLoS Med, 2007. 4(1): p. e40.
17 Christenbery, T.L., Manuscript peer review: a guide for advanced practice nurses. J Am Acad Nurse Pract, 2011. 23(1): p. 15-22.
18 Clark, R.K., Peer review: a view based on recent experience as an author and reviewer. Br Dent J, 2012. 213(4): p. 153-4.
19 Costantino, G., et al., Errors in medical literature: not a question of impact. Intern Emerg Med, 2013. 8(2): p. 157-60.
20 Curzon, M.E. and P.E. Cleaton-Jones, Reviewing scientific manuscripts. Eur Arch Paediatr Dent, 2011. 12(4): p. 184-7.
21 Davidhizar, R. and G.A. Bechtel, Tips for manuscript reviewers. Nurse Author Ed, 2003. 13(3): p. 1-4.
22 Dawson, A.J. and S.M. Yentis, Contesting the science/ethics distinction in the review of clinical research. J Med Ethics, 2007. 33(3): p. 165-7.
23 De Jong, M.J., Spotting reference errors. Nurse Author Ed, 2004. 14(2): p. 7-9.
24 Duff, K., et al., On becoming a peer reviewer for a neuropsychology journal. Arch Clin Neuropsychol, 2009. 24(3): p. 201-7.
25 Eccles, M.P., et al., Implementation Science six years on--our evolving scope and common reasons for rejection without review. Implement Sci, 2012. 7: p. 71.
26 Einarson, T.R. and G. Koren, To accept or reject? A guide to peer reviewing of medical journal papers. J Popul Ther Clin Pharmacol, 2012. 19(2): p. e328-33.
27 Emerson, G.B., et al., Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. Arch Intern Med, 2010. 170(21): p. 1934-9.
28 Forster, A.J., et al., Reliability of the peer-review process for adverse event rating. PLoS One, 2012. 7(7): p. e41239.
29 Gasparyan, A.Y. and G.D. Kitas, Best peer reviewers and the quality of peer review in biomedical journals. Croat Med J, 2012. 53(4): p. 386-9.

30 Gasparyan, A.Y., et al., Writing a narrative biomedical review: considerations for authors, peer reviewers, and editors. Rheumatol Int, 2011. 31(11): p. 1409-17.
31 Gasparyan, A.Y., L. Ayvazyan, and G.D. Kitas, Authorship problems in scholarly journals: considerations for authors, peer reviewers and editors. Rheumatol Int, 2013. 33(2): p. 277-84.
32 Gollogly, L. and H. Momen, Ethical dilemmas in scientific publication: pitfalls and solutions for editors. Rev Saude Publica, 2006. 40 Spec no.: p. 24-9.
33 Grainger, D.W., Peer review as professional responsibility: a quality control system only as good as the participants. Biomaterials, 2007. 28(34): p. 5199-203.
34 Gupta, P., et al., What is submitted and what gets accepted in Indian Pediatrics: analysis of submissions, review process, decision making, and criteria for rejection. Indian Pediatr, 2006. 43(6): p. 479-89.
35 Hall, P.A., R. Poulsom, and J. Wixon, How does The Journal of Pathology deal with conflict of interest? J Pathol, 2009. 219(4): p. 396-9.
36 Harden, R.M. and P. Lilley, A fresh approach to publishing and reviewing papers in health professions education. Med Teach, 2013. 35(1): p. 1-3.
37 Harris, A., R. Reeder, and J. Hyun, Survey of editors and reviewers of high-impact psychology journals: statistical and research design problems in submitted manuscripts. J Psychol, 2011. 145(3): p. 195-209.
38 Harris, A.H., R. Reeder, and J.K. Hyun, Common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals: what editors and reviewers want authors to know. J Psychiatr Res, 2009. 43(15): p. 1231-4.
39 Henly, S.J. and M.C. Dougherty, Quality of manuscript reviews in nursing research. Nurs Outlook, 2009. 57(1): p. 18-26.
40 Henly, S.J., J.A. Bennett, and M.C. Dougherty, Scientific and statistical reviews of manuscripts submitted to Nursing Research: Comparison of completeness, quality, and usefulness. Nurs Outlook, 2010. 58(4): p. 188-99.
41 Herbison, P., How to make your article more acceptable for the statistical reviewer. Neurourol Urodyn, 2007. 26(3): p. 318-22.
42 Houry, D., S. Green, and M. Callaham, Does mentoring new peer reviewers improve review quality? A randomized trial. BMC Med Educ, 2012. 12: p. 83.

43 Hrynaszkiewicz, I., et al., Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. Trials, 2010. 11: p. 9.

44 Jacobson, A.F., K. Schmidt, and H. Coeling, Preserving blind peer review of electronic manuscript files. Nurse Author Ed, 2005. 15(1): p. 1-4, 7.

45 Johnston, S.C., et al., Early editorial manuscript screening versus obligate peer review: a randomized trial. Ann Neurol, 2007. 61(4): p. A10-2.

46 Jones, A.W., The distribution of forensic journals, reflections on authorship practices, peer-review and role of the impact factor. Forensic Sci Int, 2007. 165(2-3): p. 115-28.

47 Kadar, N., Systemic bias in peer review: suggested causes, potential remedies. J Laparoendosc Adv Surg Tech A, 2010. 20(2): p. 123-8.

48 Kearney, M.H. and M.C. Freda, Nurse editors' views on the peer review process. Res Nurs Health, 2005. 28(6): p. 444-52.

49 Kliewer, M.A., et al., Reviewing the reviewers: comparison of review quality and reviewer characteristics at the American Journal of Roentgenology. AJR Am J Roentgenol, 2005. 184(6): p. 1731-5.

50 Landkroon, A.P., et al., Quality assessment of reviewers' reports using a simple instrument. Obstet Gynecol, 2006. 108(4): p. 979-85.

51 Landy, D.C., et al., Prepublication review of medical ethics research: cause for concern. Acad Med, 2009. 84(4): p. 495-7.

52 Lane, J.A., et al., A Peer Review Intervention for Monitoring and Evaluating sites (PRIME) that improved randomized controlled trial conduct and performance. J Clin Epidemiol, 2011. 64(6): p. 628-36.

53 Lee, K.P., et al., Predictors of publication: characteristics of submitted manuscripts associated with acceptance at major biomedical journals. Med J Aust, 2006. 184(12): p. 621-6.

54 Leek, J.T., M.A. Taub, and F.J. Pineda, Cooperation between referees and authors increases peer review accuracy. PLoS One, 2011. 6(11): p. e26895.

55 Levine, A.M., J.D. Heckman, and R.N. Hensinger, The art and science of reviewing manuscripts for orthopaedic journals: Part I. Defining the review. Instr Course Lect, 2004. 53: p. 679-88.

56 Levine, A.M., J.D. Heckman, and R.N. Hensinger, The art and science of reviewing manuscripts for orthopaedic journals: Part II. Optimizing the manuscript: practical hints for improving the quality of reviews. Instr Course Lect, 2004. 53: p. 689-97.

57 Lippert, S., M.L. Callaham, and B. Lo, Perceptions of conflict of interest disclosures among peer reviewers. PLoS One, 2011. 6(11): p. e26900.

58 Loonen, M.P., J.J. Hage, and M. Kon, Who benefits from peer review? An analysis of the outcome of 100 requests for review by Plastic and Reconstructive Surgery. Plast Reconstr Surg, 2005. 116(5): p. 1461-72; discussion 1473-5.

59 Lovejoy, T.I., T.A. Revenson, and C.R. France, Reviewing manuscripts for peer-review journals: a primer for novice and seasoned reviewers. Ann Behav Med, 2011. 42(1): p. 1-13.

60 Ludbrook, J., Peer review of biomedical manuscripts: an update. J Clin Neurosci, 2003. 10(5): p. 540-2.

61 Mazumdar, M., S. Banerjee, and H.L. Van Epps, Improved reporting of statistical design and analysis: guidelines, education, and editorial policies. Methods Mol Biol, 2010. 620: p. 563-98.

62 McLeod, P., et al., Peer review: an effective approach to cultivating lecturing virtuosity. Med Teach, 2013. 35(4): p. e1046-51.

63 Morton, J.P., Reviewing scientific manuscripts: how much statistical knowledge should a reviewer really know? Adv Physiol Educ, 2009. 33(1): p. 7-9.

64 Nahata, M.C., Tips for writing and publishing an article. Ann Pharmacother, 2008. 42(2): p. 273-7.

65 Nelson, C.A., S.R. Freeman, and R.P. Dellavalle, Reviewing dermatology manuscripts and publications. Dermatol Clin, 2009. 27(2): p. 201-4, viii.

66 Newton, D.P., Quality and peer review of research: an adjudicating role for editors. Account Res, 2010. 17(3): p. 130-45.

67 Peh, W.C. and K.H. Ng, Role of the manuscript reviewer. Singapore Med J, 2009. 50(10): p. 931-3; quiz 934.

68 Pierson, D.J., The top 10 reasons why manuscripts are not accepted for publication. Respir Care, 2004. 49(10): p. 1246-52.

69 Provenzale, J.M. and R.J. Stanley, A systematic guide to reviewing a manuscript. AJR Am J Roentgenol, 2005. 185(4): p. 848-54.

70 Rajesh, A., G. Cloud, and M.G. Harisinghani, Improving the quality of manuscript reviews: impact of introducing a structured electronic template to submit reviews. AJR Am J Roentgenol, 2013. 200(1): p. 20-3.

71 Ramulu, V.G., et al., Development of a case report review instrument. Int J Clin Pract, 2005. 59(4): p. 457-61.

72 Rangachari, P.K., Teaching undergraduates the process of peer review: learning by doing. Adv Physiol Educ, 2010. 34(3): p. 137-44.

73 Rivara, F.P., et al., A comparison of reviewers selected by editors and reviewers suggested by authors. J Pediatr, 2007. 151(2): p. 202-5.

74 Rowe, B.H., et al., Reviewer agreement trends from four years of electronic submissions of conference abstract. BMC Med Res Methodol, 2006. 6: p. 14.

75 Rutkowski, J.L. and J.V. Cairone, How to review scientific manuscripts and clinical case reports for Journal of Oral Implantology. J Oral Implantol, 2009. 35(6): p. 310-4.

76 Schroter, S., et al., Effects of training on quality of peer review: randomised controlled trial. BMJ, 2004. 328(7441): p. 673.

77 Schroter, S., et al., What errors do peer reviewers detect, and does training improve their ability to detect them? J R Soc Med, 2008. 101(10): p. 507-14.

78 Shattell, M.M., et al., Authors' and editors' perspectives on peer review quality in three scholarly nursing journals. J Nurs Scholarsh, 2010. 42(1): p. 58-65.

79 Sosa, J.A., et al., Evaluating the surgery literature: can standardizing peer-review today predict manuscript impact tomorrow? Ann Surg, 2009. 250(1): p. 152-8.80 Steinhauser, G., et al., Peer review versus editorial review and their role in innovative science. Theor Med Bioeth, 2012. 33(5): p. 359-76.

81 Timmer, A., L.R. Sutherland, and R.J. Hilsden, Development and evaluation of a quality score for abstracts. BMC Med Res Methodol, 2003. 3: p. 2.

82 Tite, L. and S. Schroter, Why do peer reviewers decline to review? A survey. J Epidemiol Community Health, 2007. 61(1): p. 9-12.

83 Triggle, C.R. and D.J. Triggle, What is the future of peer review? Why is there fraud in science? Is plagiarism out of control? Why do scientists do bad things? Is it all a case of: "all that is necessary for the triumph of evil is that good men do nothing"? Vasc Health Risk Manag, 2007. 3(1): p. 39-53.

84 Vinther, S., et al., Same review quality in open versus blinded peer review in "Ugeskrift for Laeger". Dan Med J, 2012. 59(8): p. A4479.

85 Wager, E., E.C. Parkin, and P.S. Tamber, Are reviewers suggested by authors as good as those chosen by editors? Results of a rater-blinded, retrospective study. BMC Med, 2006. 4: p. 13.

86 Wagner, A.K., et al., Peer review: issues in physical medicine and rehabilitation. Am J Phys Med Rehabil, 2003. 82(10): p. 790-802.

87 Weil, J., Peer review: an essential step in the publishing process. J Genet Couns, 2004. 13(3): p. 183-7

88 Williams, E.H., P.A. Carpentier, and T. Misteli, Minimizing the "re" in review. J Cell Biol, 2012. 197(3): p. 345-6.

89 Wong, V.S., The examination of peer review and publication in neurology. J Child Neurol, 2010. 25(10): p. 1298-301.

90 Yaffe, M.B., Re-reviewing peer review. Sci Signal, 2009. 2(85): p. eg11.

# 4. Les interventions visant à améliorer la qualité de l'évaluation par les peer-reviewers des revues biomédicales

## 4.1 Justification de l'étude

Malgré un intérêt croissant, les fonctions des peer-reviewers sont mal définies (15) et l'impact même de la peer-review pose question (85) (86) (82). D'autant plus que certaines études ont mis en évidence que les peer-reviewers ne sont pas à même de détecter des erreurs méthodologiques (35), d'améliorer la manière dont sont rapportées les études (36) ou de diminuer les erreurs dans la présentation des résultats (87) (88) (32). Dans le but d'améliorer la qualité de la peer-review, certains éditeurs ont développé des interventions (89) (90) (91). En 2007, Jefferson et al. (1) ont publié une revue systématique de la littérature de ses interventions dans la peer-review éditoriale. Les auteurs avaient inclus les études rétrospectives et prospectives et avaient conclu qu'il n'y avait que peu d'évidence de la peer-review éditoriale dans l'amélioration de la qualité des manuscrits. A notre connaissance, il n'existait de revue systématique récente.

## 4.2 Objectif

L'objectif de notre recherche était d'identifier les interventions visant à améliorer la peer-review et ayant démontré leur efficacité.

## 4.3 Méthodes

Nous avons réalisé une recherche sur différentes bases de données : *Cochrane Central Register of Controlled Trials* (CENTRAL; June 2015, Issue 6), *MEDLINE via PubMed*, Embase, et *WHO International Clinical Trials Registry Platform*, pour tous les essais contrôlés randomisés évaluant une intervention visant à améliorer la qualité de la peer-review dans les journaux biomédicaux (92). Nous avons inclus les essais contrôlés randomisés publiés en anglais, quelle

que soit leur unité de randomisation (i.e., manuscrit ou peer-reviewer). Nous avons exclu les essais contrôlés randomisés ayant évalué la présence et/ou l'effet du peer-reviewer sur le critère de jugement, les doublons et les essais non randomisés.

Nous avions pré spécifié les différentes interventions possibles : 1) entraînement/formation ; 2) l'ajout d'un peer-reviewer spécifique (i.e., statisticien), 3) l'utilisation de checklist, 4) évaluation en ouvert, 5) évaluation en aveugle ou masquée, 6) autres interventions.

Les critères de jugement étaient les suivants : 1) qualité finale du manuscrit (93) (94) 2) qualité du rapport du peer-reviewer (95) (96) 3) le taux de rejet 4) le temps de la durée de l'évaluation par le peer-reviewer 5) la durée totale du processus de peer-review.

L'extraction des données a été réalisée indépendamment par 2 chercheurs en utilisant un formulaire d'extraction standardisé. Le risque de biais de chaque essai inclus a été évalué selon le *Handbook* de la Collaboration Cochrane.

## 4.4 Résultats

Nous avons sélectionné 22 rapports d'essais contrôlés randomisés évaluant 25 comparaisons : cinq d'interventions de formation (97) (98) ((99)((100), deux sur l'ajout d'un évaluateur statistique ((101) (102) deux sur l'utilisation d'une checklist (103) ((104), sept sur l'examen par les peer-reviewers en ouvert (i.e., peer-reviewers informés que leur identité serait révélée) (105) (106) (107) (108) (109), six sur l'évaluation en aveugle par les peer-reviewers (i.e., peer-reviewers ignoraient les noms et les affiliations des auteurs)(110) (111) (112) et trois autres interventions pour accélérer le processus d'évaluation par les peer-reviewers ((113) (114) (115). Seuls sept essais randomisés ont été publiés depuis 2004. Par rapport au processus standard d'évaluation par les peer-reviewers, la formation n'a pas amélioré la qualité du rapport

d'évaluation des peer-reviewers et l'utilisation d'une checklist n'a pas amélioré la qualité du manuscrit final. L'ajout d'un examen statistique spécifique a amélioré la qualité du manuscrit final (différence moyenne standardisée (SMD), 0,58, IC à 95%, 0,19 à 0,98). L'évaluation en ouvert a amélioré la qualité du rapport des peer-reviewers (SMD, 0,14, IC à 95%, 0,05 à 0,24), n'a pas affecté le temps passé par ceux-ci à réaliser la peer-review (différence moyenne : 0,18; 0,43), mais a diminué le taux de rejet (Odds Ratio, 0,56; IC 95%, 0,33 à 0,94). L'examen en aveugle par les peer-reviewers n'a pas affecté la qualité du rapport d'évaluation des peer-reviewers ni le taux de rejet. Les interventions visant à accélérer le processus de peer-review étaient trop hétérogènes pour permettre la mise en commun des résultats.

## 4.5 Discussion et limites

Malgré le rôle essentiel de la peer-review dans la recherche biomédicale, seuls 22 essais contrôlés randomisés ont évalué des interventions pour améliorer sa qualité, et seules 11 études ont été réalisées au cours des dix dernières années. Le premier congrès dédié à la peer-review en 1989 à Chicago avait mis l'accent sur l'importance de stimuler la recherche dans ce domaine (116). Plusieurs années plus tard, malgré quelques initiatives comme l'action Européenne COST PEERE (117), la recherche impliquant un design expérimental est rare dans ce domaine, ce qui est alarmant compte tenu du coût et du rôle central de ce processus en science biomédicale.

Les difficultés liées à la recherche sur la peer-review sont multiples et complexes. Premièrement, le rôle du peer-review n'est pas consensuel (15) (118). Deuxièmement, le design de l'étude pose question : randomisation des peer-reviewers ou des journaux ? Un même essai évalué par tous les participants ou un seul essai par participant ? Troisièmement, le biais induit par la gestion de l'aveugle dans ces essais randomisés est problématique. En effet, par exemple lors de l'évaluation

des peer-reviewers en ouvert, les participants sont obligatoirement informés de l'intervention et donc cela peut induire un biais de recrutement.

Enfin, le choix du critère de jugement principal est très compliqué dans ce domaine. Les critères les plus souvent utilisés sont soit une évaluation subjective de la qualité de la peer-review par les éditeurs, soit l'évaluation de la qualité du rapport des peer-reviewers, soit la qualité du manuscrit final.

Notre étude présente des limites. Premièrement, nous nous sommes intéressés uniquement aux essais randomisés dans le domaine biomédical. De plus, le peu d'études identifiées, leurs faibles effectifs, leur qualité méthodologique ainsi que leur applicabilité, limite l'interprétation de nos résultats.

## 4.6 Conclusion

Malgré le rôle essentiel des peer-reviewers seules quelques interventions ont été évaluées dans des essais contrôlés randomisés. La recherche dans ce domaine être développée dans des revues biomédicales.

## 4.7 Article

Bruce et al. BMC Medicine (2016) 14:85
DOI 10.1186/s12916-016-0631-5

BMC Medicine

RESEARCH ARTICLE Open Access

# Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis

Rachel Bruce[1,2,3†], Anthony Chauvin[2,3,4†], Ludovic Trinquart[2,3,5], Philippe Ravaud[1,2,3,5] and Isabelle Boutron[2,3,5*]

**Abstract**

**Background:** The peer review process is a cornerstone of biomedical research. We aimed to evaluate the impact of interventions to improve the quality of peer review for biomedical publications.

**Methods:** We performed a systematic review and meta-analysis. We searched CENTRAL, MEDLINE (PubMed), Embase, Cochrane Database of Systematic Reviews, and WHO ICTRP databases, for all randomized controlled trials (RCTs) evaluating the impact of interventions to improve the quality of peer review for biomedical publications.

**Results:** We selected 22 reports of randomized controlled trials, for 25 comparisons evaluating training interventions (n = 5), the addition of a statistical peer reviewer (n = 2), use of a checklist (n = 2), open peer review (i.e., peer reviewers informed that their identity would be revealed; n = 7), blinded peer review (i.e., peer reviewers blinded to author names and affiliation; n = 6) and other interventions to increase the speed of the peer review process (n = 3). Results from only seven RCTs were published since 2004. As compared with the standard peer review process, training did not improve the quality of the peer review report and use of a checklist did not improve the quality of the final manuscript. Adding a statistical peer review improved the quality of the final manuscript (standardized mean difference (SMD), 0.58; 95 % CI, 0.19 to 0.98). Open peer review improved the quality of the peer review report (SMD, 0.14; 95 % CI, 0.05 to 0.24), did not affect the time peer reviewers spent on the peer review (mean difference, 0.18; 95 % CI, −0.06 to 0.43), and decreased the rate of rejection (odds ratio, 0.56; 95 % CI, 0.33 to 0.94). Blinded peer review did not affect the quality of the peer review report or rejection rate. Interventions to increase the speed of the peer review process were too heterogeneous to allow for pooling the results.

**Conclusion:** Despite the essential role of peer review, only a few interventions have been assessed in randomized controlled trials. Evidence-based peer review needs to be developed in biomedical journals.

**Keywords:** Peer review process, Peer reviewers, Systematic review, Meta-analysis

* Correspondence: isabelle.boutron@htd.aphp.fr
†Equal contributors
[2]INSERM, UMR 1153, Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité – (CRESS), METHODS team, Paris, France
[3]Paris Descartes University, Paris, France
Full list of author information is available at the end of the article

© 2016 The Author(s). **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

## Background

The peer review process is a cornerstone to improve the quality of scientific publications [1–3]. It is used by most scientific journals to inform editors' decisions and to improve the quality of published reports [4]. Worldwide, peer review costs an estimated £1.9 billion annually and accounts for about one-quarter of the overall costs of scholarly publishing and distribution [5, 6]. Despite this huge investment, primary functions of peer reviewers are poorly defined [7] and the impact and benefit of peer review are increasingly questioned [8–13]. Particularly, studies have shown that peer reviewers were not able to appropriately detect errors [14, 15], improve the completeness of reporting [16], or decrease the distortion of the study results [17–19]. Some interventions have been developed and implemented by some editors to improve the quality of peer review [4, 20–22]. In 2007, Jefferson et al. [23] published a systematic review evaluating the effect of processes in editorial peer review through a search performed in 2005. The authors included prospective and retrospective comparative studies and concluded that little empirical evidence was available to support the use of editorial peer review as a mechanism to improve quality in biomedical publications. To our knowledge, no recent systematic review, including studies published over the last 10 years, has been published.

We aimed to perform a systematic review and a meta-analysis of all randomized controlled trials (RCTs) evaluating the impact of interventions to improve the quality of peer review in biomedical journals.

## Methods

### Search strategy

We searched the Cochrane Central Register of Controlled Trials (CENTRAL; June 2015, Issue 6), MEDLINE (via PubMed), Embase, and the WHO International Clinical Trials Registry Platform for all reports of RCTs evaluating the impact of interventions aiming to improve the quality of peer review in biomedical journals (last search: June 15, 2015). Biomedical research is defined as the area of science devoted to the study of the processes of life, the prevention and treatment of disease, and the genetic and environmental factors related to disease and health. We also searched the Cochrane Database of Systematic Reviews to identify systematic reviews on the peer review process. We had no limitation on language or date of publication. Our search strategy relied on the Cochrane Highly Sensitive Search Strategies [24] and the search term "peer review". We hand-searched reference lists of reports and reviews dedicated to the peer review process identified during the screening process.

### Study selection

Two researchers (RB, AC) independently screened all citations retrieved by using the Resyweb software. We obtained and independently examined the full-text article for all citations selected for possible inclusion. Any disagreements were discussed with a third researcher until consensus was reached.

We included RCTs, whatever the unit of randomization (manuscript or peer reviewers), evaluating any interventions aimed at improving the quality of peer review for biomedical publications regardless of publication language. We excluded RCTs evaluating the presence and effect of peer reviewer bias on the outcome, such as positive outcome bias. Non-randomized studies were excluded. Duplicate publications of the same study were collated for each unique trial.

### Interventions

We pre-specified the categorization of the interventions evaluated as follows [4]:

1) Training, which included training or mentoring programs for peer reviewers to provide instructional support for appropriately evaluating important components of manuscript submissions. These interventions directly target the ability of peer reviewers to appropriately evaluate the quality of the manuscripts.
2) Addition of peer reviewers for specific tasks or with specific expertise such as adding a statistical peer reviewer, whose main task is to detect the misuse of methods or misreporting of statistical analyses.
3) Peer reviewers' use of a checklist, such as reporting guideline checklists, to evaluate the quality of the manuscript.
4) "Open" peer review process, whereby peer reviewers are informed that their name would be revealed to the authors, other peer reviewers, and/or the public. The purpose of these approaches is to increase transparency and thus increase the accountability of existing peer reviewers to produce good-quality peer reviews.
5) "Blinded"/masked peer review, whereby peer reviewers are blinded to author names and affiliation. Author names and/or potentially identifying credentials are removed from manuscripts sent for peer review so as to remove or minimize peer reviewer biases that arise from knowledge of and assumptions about author identities.
6) Other interventions. Any other types of interventions identified were secondarily classified.

### Outcomes

The outcomes were as follows:

1) Final manuscript quality (i.e., revised manuscript after all peer review assessments and rounds). Several scales were used to assess manuscript quality. The Manuscript Quality Assessment Instrument, a 34-item scale, each item scored from 1 to 5 [25], aimed to evaluate the quality of the research report (i.e., whether the authors have described their research in enough details and with sufficient clarity so a reader could make an independent judgment about the strength and weakness of their data and conclusion), not the quality of the research itself. This scale is organized along the same dimensions of a journal article (Title and Abstract, Introduction, Methods, Results, Discussion, Conclusions, and General evaluation). Other scales consisted of scales measuring the completeness of reporting [26] based on reporting guideline items (e.g., CONSORT items). A high score indicates a high quality manuscript.
2) Quality of the peer review report, measured by scales such as the Review Quality Instrument [27] or editor routine quality rating scales [28]. A high score indicates a high quality peer review report. Various versions of the Review Quality Instrument exist [27, 29]. The last version of the scale is based on eight items: importance, originality, methodology, presentation, constructiveness of comments, substantiation of comments and interpretation of results, and a global item. Each item is scored on a 5-point scale. This version had a high level of agreement (K = 0.83) and a good inter-rater reliability (K = 0.31) in the mean total score. No evidence of floor or ceiling effects was found.
3) Rejection rate (i.e., recommendation by peer reviewers about whether or not the reviewed manuscripts should be rejected from publication in the journal).
4) Time spent on the peer review as reported by peer reviewers. The time spent is important to provide an evaluation of the burden and possible cost of the peer review process.
5) Overall duration of the peer review process.

All outcomes were pre-specified except the time spent on the peer review, which was added after inspection of the studies revealed RCTs evaluating interventions to improve the speed of the peer review process.

### Data extraction

Data extraction was performed independently by two researchers who used a standardized data extraction form (Additional file 1). When assessment differed, the item was discussed and consensus was reached. When needed, a third reviewer assessed the report to achieve consensus.

General characteristics of the study recorded included the journal publication, study design, unit of randomization (peer reviewer or manuscripts or both), eligibility criteria for peer reviewers and for manuscripts, number of peer reviewers and/or manuscripts randomized and analyzed, and interventions compared.

The risk of bias within each RCT was assessed by the following domains of the Cochrane Collaboration risk of bias tool [30]: selection bias (methods for random sequence generation and allocation concealment), detection bias (blinding of outcome assessors) and attrition bias (incomplete outcome data). We did not assess performance bias (blinding of participants and providers) because blinding was never possible in this context. Each domain was rated as low, high or unclear risk of bias, according to the guidelines [30]. As recommended by the Cochrane Collaboration, the blinding of outcome assessors and incomplete outcome data domains were assessed at the outcome level. When the risk of bias was unclear, we contacted the authors by email to request more information. When peer reviewers participating in the RCTs were not aware of the study hypothesis or were not aware of the outcome assessed, we considered the detection bias at low risk.

We independently recorded data for each outcome. To avoid selective inclusion of results in systematic reviews, when several results were reported for the selected outcomes (e.g., several scales used to assess the manuscript quality or several time points used to assess the quality of the peer review), we retrieved the results for the outcome reported as the primary outcome in the manuscript or the final or latest time of assessment of the outcome.

For continuous outcomes (i.e., the quality score of the final manuscript, quality of the peer review report, time peer reviewers spent on the peer review), we recorded the mean and standard deviation per arm. If data were not provided in the form of a mean and standard deviation, we derived these from the reported data and produced statistics following the Cochrane handbook recommendations [31].

For dichotomous outcomes (i.e., rejection rate), we recorded the number of manuscripts evaluated and the number of manuscripts for which peer reviewers recommended rejection.

For studies with multiple intervention groups (e.g., blinded peer review and open peer review), we followed the Cochrane recommendations [31] and, when appropriate, we combined all relevant intervention groups for the study into a single group, combined all relevant control intervention groups into a single control group, or selected one pair of interventions and excluded the others.

When required, we extracted the outcome data from the published figures by using Digitizelt [32]. We contacted authors when key data were missing.

Publication bias and small study effect were assessed by visual inspection of funnel plots when appropriate.

### Data synthesis

Treatment effect measures were odds ratios (ORs) for binary outcomes (rejection rate) and standardized mean differences (SMDs) for continuous outcomes (quality score of the final manuscript, quality of peer review report, and mean difference for the time peer reviewers spent on the peer review). RCTs measured the quality of the manuscript and the peer review in various ways (using different scoring systems), so we measured the intervention effects by using SMDs.

Because of the diversity of the interventions and settings, we used random-effects meta-analysis (Dersimonian and Laird model). We assessed heterogeneity by visual inspection of the forest plots, Cochrane's homogeneity test, and the $I^2$ statistic [33]. In the meta-analysis of RCTs assessing open peer review, we performed one subgroup analysis to evaluate the impact of different intensities of open peer review (open to other peer reviewers/open to authors/publicly available). Analyses involved use of Review Manager 5.3 (Cochrane Collaboration).

RCTs assessing training interventions reported a mixture of change from baseline and final value scores regarding the quality of the peer review report. As recommended, we did not combine the final value and change scores together as SMDs [34].

## Results

### Study selection and characteristics of included studies

Figure 1 provides the flow diagram for the study selection. From the 4592 citations retrieved, we selected 21 published articles reporting 22 RCTs and 25 comparisons, evaluating training interventions (n = 5), the addition of a statistical peer reviewer (n = 2), use of a checklist (n = 2), open peer review (n = 7), blinded peer review (n = 6), and other interventions to increase the speed of the peer review process (n = 3). Only seven RCTs were published since 2004. Searching the WHO trial register platform did not identify any RCT.

The characteristics of the selected studies are presented in Table 1. The description of the interventions assessed in each RCT is provided in Additional file 2.

The 22 RCTs were published in 12 different journals. The study design varied according to the interventions evaluated and consisted of randomizing peer reviewers (n = 15), manuscripts (n = 6) or both (n = 4). The manuscripts used in the RCTs were submitted by authors to the system by the usual process (n = 20) or were fictitious manuscripts specifically prepared for the study with error added (n = 2).

### Data synthesis

#### *Training*

Four RCTs, involving five comparisons and 616 peer reviewers, compared the impact of training interventions to the standard peer review procedure for the given journal on the quality of the peer review report [35–38]. Four comparisons of the *Annals of Emergency Medicine* [36–38] evaluated feedback (n = 2), training (n = 1) and mentoring (n = 1), and one comparison of the *British Medical Journal* evaluated training [35]. Different kinds of training were evaluated: structured workshops with journal editors or self-taught training with a package created especially for peer reviewers. Mentoring consisted of new peer reviewers discussing their review with a senior peer reviewer before sending the review. Feedback consisted of peer reviewers receiving a copy of the editor's rating of their review. One study evaluated two kinds of interventions: self-taught or face-to-face [35]. For this study, we pooled these two arms versus the usual process. The risk of attrition bias was high for four RCTs (Additional file 3). Training had no impact on the quality of the peer review report, but the heterogeneity across studies was high (66.2 %) and the 95 % confidence interval (95 % CI) relatively large (Fig. 2). Only one study assessed the time to perform the peer review and the rejection rate with training [35]: training did not affect the time to review manuscripts (mean (SD), 116.9 (69.4) min for training vs. 108.5 (70.5) min for the usual process; $P = 0.38$) – but did increase recommendations to reject manuscripts (OR, 0.47; 95 % CI, 0.42 to 0.78).

#### *Addition of a statistical peer reviewer*

Two RCTs examined the impact of adding a statistical peer reviewer on the final manuscript quality [39, 40]. All RCTs involved the *Medicina Clínica* journal (a Spanish journal of internal medicine) and were conducted by the same research team. The RCT by Cobo et al. [40] was a $2 \times 2$ factorial design comparing the addition of a statistical peer reviewer, the use of a checklist, both interventions, and the usual process. For this analysis, we used the arms dedicated to the usual process and the addition of a statistical peer reviewer. The risk of bias was low for these two studies (Additional file 3). Overall, 105 manuscripts were included in the meta-analysis. Adding a statistical peer reviewer increased the quality of the final manuscript as compared to the usual process (combined SMD, 0.58; 95 % CI, 0.19 to 0.98; $I^2 = 0$ %; Fig. 3).

#### *Use of a reporting guideline checklist*

Two studies evaluated the impact of peer reviewers using a reporting guideline checklist on the quality of

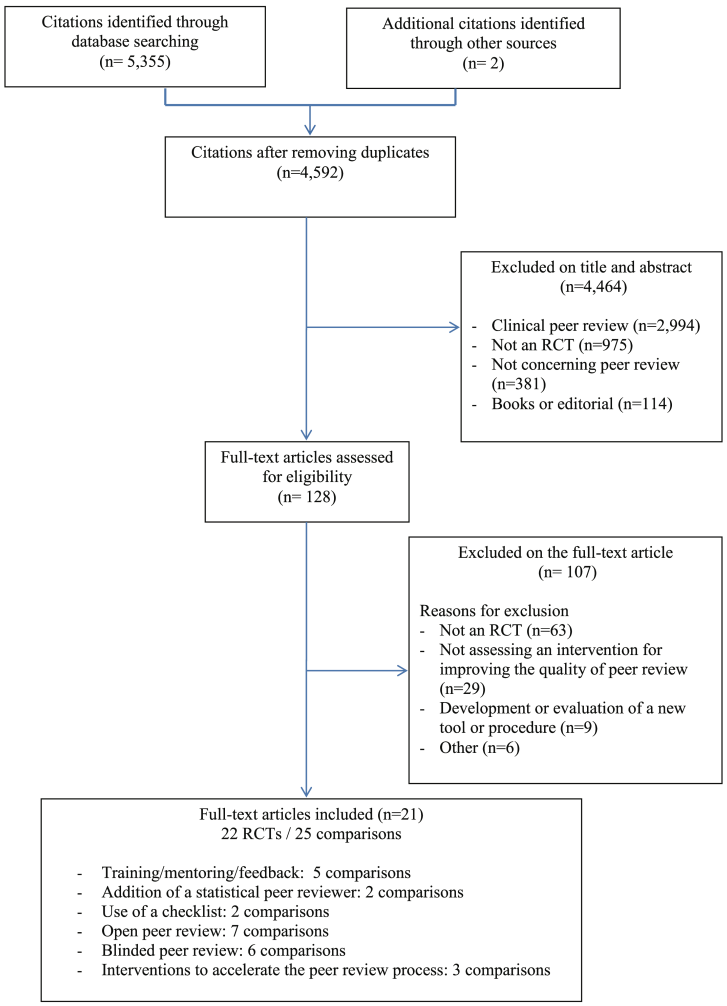



**Fig. 1** Study selection flow diagram

the final manuscript [40, 41]. These studies were performed by the same research team and involved the *Medicina Clínica* journal. For the RCT by Cobo [40], we used the arms dedicated to the usual process and the addition of a checklist. The risk of bias was low (Additional file 3). Overall, 152 manuscripts were included in the meta-analysis. Overall, use of a checklist had no effect on the quality of the final manuscript as compared with the usual process (combined SMD, 0.19; 95 % CI, −0.22 to 0.61; $I^2$ = 38 %; Fig. 4).

### *"Open" peer review*

Seven RCTs assessed the impact of open review interventions [42–48], consisting of informing peer reviewers that their identity would be revealed to other peer reviewers only (n = 2), to the authors of the manuscript being reviewed (n = 3), and to the general public via web-posting (n = 1). One study evaluated both open peer review and peer reviewers blinded or not to the authors' identity [47]. One study was not included in the meta-analysis because the usual process was not the

**Table 1** General characteristics of the trials included

| Author, Year, Journal publishing the study | Intervention/Comparator | Journal(s) involved in the study | Peer reviewers (n) | Manuscripts (n)/Peer review reports (n) | Unit of randomization | Outcomes (Scale) | Sample size (randomized/ analyzed) | Risk of bias |
|---|---|---|---|---|---|---|---|---|
| Training/Mentoring/Feedback | | | | | | | | |
| Callaham, JAMA 2002 Study 1a [36] | Feedback by editors/usual process | Annals of Emergency Medicine | Low volume low quality[a] peer reviewers (n = 51) | Manuscripts submitted to the journal (n = NR)/182 peer review reports | Peer reviewers | The quality of peer review report (using editor routine quality scale) | 51 randomized/35 analyzed | - Rs: Low<br>- AI: Low<br>- D: Low<br>- At: High |
| Callaham, JAMA 2002 Study 1b [36] | Feedback by editors/usual process | Annals of Emergency Medicine | Low volume, average quality[b] peer reviewers (n = 127) | Manuscripts submitted to the journal (n = NR)/324 peer review reports | Peer reviewers | The quality of peer review report (using editor routine quality scale) | 127 randomized/ 95 analyzed | - Rs: Low<br>- AI: Low<br>- D: Low<br>- At: High |
| Callaham, 2002 Ann Emerg Med Study 2 [37] | Feedback by editors/usual process | Annals of Emergency Medicine | Average quality[b] peer reviewers (n = 150) | Manuscripts submitted to the journal (n = NR)/79 peer review reports | Peer reviewers | The quality of peer review report (using editor routine quality scale) | 150 randomized/ 22 analyzed | - Rs: Low<br>- AI: Low<br>- D: Low<br>- At: High |
| Houry, 2012 BMC Med Educ [38] | Training (workshop)/usual process | Annals of Emergency Medicine | New peer reviewers (n = 50) | Manuscripts submitted to the journal (n = 490)/490 peer review reports | Peer reviewers | The quality of peer review report (using editor routine quality scale) | 50 randomized/46 analyzed | - Rs: Low<br>- AI: Unclear<br>- D: Unclear<br>- At: Low |
| Schroter, 2004 BMJ [35] | Training (face-to-face or self-training)/usual process (The two intervention groups, face-to-face and self-training, were pooled in the meta-analysis) | British Medical Journal | Consenting peer reviewers (n = 609) | One fabricated manuscript with errors (n = 1)[c]/418 peer review reports | Peer reviewers | 1) The quality of peer review report (using the RQI Version 3.2)<br>2) The rejection rate<br>3) The time spent on the review | 609 randomized/ 418 analyzed | - Rs: Low<br>- AI: Low<br>- D: Low<br>- At: High |
| Statistical peer review | | | | | | | | |
| Arnau, 2003 Med Clin (Barc) [39] | Adding a statistical peer reviewer/ usual process | Medicina Clinica | Statistical peer reviewers (n = NR) | Manuscripts submitted to the journal (n = 82) | Manuscripts | The final manuscript quality (using the MQAI) | 82 randomized/43 analyzed | - Rs: Low<br>- AI: Low<br>- D: Low<br>- At: High |
| Cobo, 2007 PLOS One [40] | 2 × 2 factorial design comparing/ adding a statistical peer reviewer/ use of a reporting guidelines checklist/both/usual process (we selected only the two groups | Medicina Clinica | Statistical peer reviewers (n = 39) | Consecutive manuscripts submitted to the journal (n = 68) | Manuscripts | The final manuscript quality (using the MQAI) | 68 randomized/62 analyzed | - Rs: Low<br>- AI: Low<br>- D: Low<br>- At: Low |

**Table 1** General characteristics of the trials included *(Continued)*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | adding a statistical peer reviewer/ usual process in the analysis)[d] | | | | | | | |
| Checklist | | | | | | | | |
| Cobo, 2007 PLOS One [40] | 2 × 2 factorial design comparing adding a statistical peer reviewer/ use of a reporting guidelines checklist/both/usual process (we selected only the two groups use of a reporting guidelines checklist/usual process in the analysis)[d] | Medicina Clinica | Statistical peer reviewers (n = 39) | Consecutive manuscripts submitted to the journal (n = 69) | Manuscripts | The final manuscript quality (using the MQAI) | 69 randomized/60 analyzed | - Rs: Low<br>- Al: Low<br>- D: Low<br>- At: Low |
| Cobo, 2011 BMJ [41] | Use of a reporting guidelines checklist/usual process | Medicina Clinica | A senior statistician (n = 1) | Consecutive manuscripts submitted to the journal (n = 92) | Manuscripts | The final manuscript quality (using the MQAI) | 92 randomized/92 analyzed | - Rs: Low<br>- Al: Low<br>- D: Low<br>- At: Low |
| Open peer review (i.e., identity of peer reviewers revealed to the authors, other peer reviewers, and/or general public) | | | | | | | | |
| Das Sinha, 1999 Natl Med J India [42] | Pairs of reviewers were identified to assess each manuscript and the two reviewers were randomized; one to be informed they would have their identity revealed to the other peer reviewer and one to remain anonymous | The National Medical Journal of India | Peer reviewers of the journal (n = 156, 78 pairs) | Manuscripts submitted to the journal (n = 100)/156 peer review reports | Pairs of peer reviewers | 1) The quality of peer review report (using editor routine quality scale)<br>2) The rejection rate | 100 randomized/ 78 analyzed | - Rs: Low<br>- Al: Low<br>- D: Low<br>- At: Low |
| Godlee, 1998 JAMA [47] | Five groups: 1) ask to sign their report + blinded to authors name and affiliation; 2) ask to remain anonymous + blinded to authors name and affiliation; 3) ask to sign their report + unblinded to authors name and affiliation; 4) ask to remain anonymous + unblinded to authors name and affiliation/ usual process<br>We pooled groups 1 and 3 vs. 2 and 4<br>A fifth group where peer reviewers were unaware of the study for which the manuscript was sent according to the usual process was not taken into account in the analysis | British Medical Journal | Peer reviewers of the journal (n = 420) | One fabricated manuscript with errors (n = 1)/184 peer review reports | Peer reviewers | The rejection rate | 360 randomized/ 184 analyzed | - Rs: Low<br>- Al: Low<br>- D: Low<br>- At: High |
| Van Rooyen, 1998 JAMA [48] | Open to peer reviewers (combination of blinded and unblinded to authors identity) vs. anonymous (combination | British Medical Journal | Peer reviewers of the journal (n = NR) | Consecutive manuscripts submitted to the journal (n = 527) | Manuscripts & peer reviewers | 1) The quality of peer review report (using | 527 manuscript randomized/598 reviews analyzed | - Rs: Low<br>- Al: Unclear<br>- D: Low |

**Table 1** General characteristics of the trials included *(Continued)*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | of blinded and unblinded to authors identity) | | | | | the RQI Version 3.2)<br>2) The time spent on the review | | - At: Unclear |
| Van Rooyen, 1999 BMJ [43] | Identity revealed to authors/ peer reviewers remained anonymous to authors<br>For each manuscript, a pair of reviewers were identified and each reviewer was randomized to have their identity revealed to authors or remain anonymous | British Medical Journal | Peer reviewers of the journal (n = 250) | Consecutive manuscripts submitted to the journal (n = 125)/113 manuscripts assessed/226 peer review reports | Peer reviewers | 1) The quality of peer review report (using the RQI Version 4)<br>2) The rejection rate<br>3) The time spent on the review | 250 randomized/ 226 analyzed | - Rs: Low<br>- AI: Unclear<br>- D: Low<br>- At: High |
| Van Rooyen, 2010 BMJ [44] | Identity revealed to general public/peer reviewers signed their review for authors and other peer reviewers | British Medical Journal | Peer reviewers of the journal (n = 471) | Consecutive manuscripts submitted to the journal (n = 558) | Manuscripts | 1) The quality of peer review report (using the RQI Version 4)<br>2) The rejection rate<br>3) The time spent on the review | 558 manuscript randomized/471 analyzed | - Rs: Unclear<br>- AI: Unclear<br>- D: Low<br>- At: Low |
| Vinther, 2012 Dan Med [45] | Pairs of reviewers were identified to assess each manuscript; for each manuscript, peer reviewers were randomized to have their identity revealed to authors/remained anonymous to authors | The Journal of Danish Medical Association | Peer reviewers of the journal (n = 380) | Manuscripts submitted to the journal (n = 190)/364 peer review reports | Peer reviewers | 1) The quality of peer review report (using the RQI Version 4)<br>2) The rejection rate | 380 randomized/ 364 analyzed | - Rs: Unclear<br>- AI: Unclear<br>- D: Low<br>- At: Low |
| Walsh, 2000 Br J Psychiatry [46] | Identity revealed to authors/ peer reviewers remained anonymous to authors | British Journal of Psychiatry | Peer reviewers of the journal (n = 245) | Manuscripts submitted to the journal (n = 408)/354 peer review reports | Manuscript | 1) The quality of peer review report (using the RQI Version 3.2)<br>2) The rejection rate<br>3) The time spent on the review | 408 manuscripts randomized/354 analyzed | - Rs: Low<br>- AI: Unclear<br>- D: Low<br>- At: High |
| Blinded peer review (i.e., peer reviewers are blinded of the authors name and affiliation) | | | | | | | | |
| Alam, 2011 Br J Dermatol [49] | Randomization of four peer reviewers for each manuscript, two randomized to assess a blinded version of the manuscript, | Dermatologic Surgery | Volunteer peer reviewers of the journal (n = 20) | Consecutive manuscripts submitted to the journal (n = 40)/160 peer review reports | Peer reviewers & manuscript | The rejection rate | 20 peer reviewers/ 40 manuscripts 160 peer review reports analyzed | - Rs: Unclear<br>- AI: Unclear<br>- D: Low |

**Table 1** General characteristics of the trials included *(Continued)*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | two to assess an unblinded version of the manuscript | | | | | | 6 peer reviewers failed to do their review and were replaced (3 blinded/3 unblinded) | - At: Low |
| Fisher, 1994 JAMA [50] | Identification of four peer reviewers for each manuscript, two randomized to assess a blinded version of the manuscript, two to assess an unblinded version of the manuscript | Journal of Developmental and Behavioral Pediatrics | Peer reviewers of the journal (n = 228) | Consecutive manuscripts submitted to the journal (n = 57)/228 peer review reports | Peer reviewers | The rejection rate | 228 randomized/ 220 analyzed | - Rs: Low - AI: Unclear - D: Low - At: Low |
| Godlee, 1998 JAMA [47] | Five groups: 1) ask to sign their report + blinded to authors name and affiliation; 2) ask to remain anonymous + blinded to authors name and affiliation; 3) ask to sign their report + unblinded to authors name and affiliation; 4) ask to remain anonymous + unblinded to authors name and affiliation/usual process<br>We pooled groups 1 and 2 vs. 3 and 4<br>A fifth group where peer reviewers were unaware of the study for which the manuscript was sent according to the usual process was not taken into account in the analysis | British Medical Journal | Peer reviewers of the journal (n = 360) | One fabricated manuscript with errors (n = 1)/130 peer review reports | Peer reviewers | The rejection rate | 240 randomized/ 130 analyzed | - Rs: Low - AI: Low - D: Low - At: High |
| Justice, 1998 JAMA [51] | Identification of two peer reviewers for each manuscript, one randomized to assess a blinded version of the manuscript, one to assess an unblinded version of the manuscript | Annals of Emergency Medicine, Annals of Internal Medicine, JAMA, Obstetrics & Gynecology and Ophthalmology | Peer reviewers of journals (n = NR) | Manuscripts submitted to journals (n = 92)/77 manuscripts with two peer review reports | Peer reviewers | The quality of peer review report (using editor routine quality scale) | 92 manuscript with two reviewers/77 manuscript with two reviewers reports analyzed | - Rs: Low - AI: Unclear - D: Low - At: Unclear |
| McNutt, 1990 JAMA [52] | Identification of two peer reviewers for each manuscript, one randomized to assess a blinded version of the manuscript, one to assess an unblinded version of the manuscript | Journal of General Internal Medicine | Peer reviewers of the journal (n = NR) | Manuscripts submitted to the journal (n = 127)/252 peer review reports | Peer reviewers | 1) The quality of peer review report (using editor routine quality scale)<br>2) The time spent on the review | 127 manuscript with two reviewers/127 and 125 reviewers reports analyzed | - Rs: Low - AI: Unclear - D: Low - At: Low |
| Van Rooyen, 1998 JAMA [48] | Identification of two peer reviewers for each manuscript, one randomized to assess a blinded version of the manuscript, one to assess an | British Medical Journal | Peer reviewers of the journal (n = NR) | Consecutive manuscripts submitted to the journal included in this study (n = 309) | Peer reviewers & manuscripts | 1) The quality of peer review report (using the RQI Version 3.2) | 309 manuscripts with two reviewers randomized/618 | - Rs: Low - AI: Unclear - D: Low |

**Table 1** General characteristics of the trials included *(Continued)*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | unblinded version of the manuscript<br>This study also assessed masked versus unmasked review (See above in open peer review intervention) | | | | | 2) The time spent on the review | peer reviews reports analyzed | - At: Unclear |
| Accelerate the peer review process | | | | | | | | |
| Johnston, 2007 Ann Neurol [54] | Early screening by editors/formal external review | Annals of Neurology | Peer reviewers of the journal (n = 386) | Consecutive manuscripts submitted to the journal (n = 351) | Manuscripts | The time to a manuscript decision | 351 manuscripts randomized/351 reviews analyzed | - Rs: Unclear<br>- Al: Low<br>- D: low<br>- At: Low |
| Neuhauser, 1989 Medical Care [55] | Calling first peer reviewers/ sending out manuscript without a prior phone call | Medical Care | Peer reviewers of the journal (n = NR) | Manuscripts submitted to the journal (n = 95) | Peer reviewers | The overall time for the peer review process | 95 manuscripts with two peer reviewers randomized | - Rs: Unclear<br>- Al: Unclear<br>- D: Low<br>- At: Low |
| Pitkin, 2002 JAMA [53] | Asking first: referees received information about manuscript by fax and indicated their willingness to review/editors mailed the manuscript and asked to return the review | Obstetrics & Gynecology | Peer reviewers of the journal (n = NR) | Consecutive manuscripts submitted to the journal (n = 283)<br>Identification of two peer reviewers for each manuscript | Peer reviewers | 1) The overall time for the peer review process<br>2) The quality of peer review report (using editor routine quality scale) | 283 manuscripts with two reviewers randomized and analyzed | - Rs: Low<br>- Al: Low<br>- D: Low<br>- At: Low |

Rs, Random sequence generation; Al, Allocation concealment; D, Detection bias (blinding of outcome assessment); At, Attrition bias (incomplete outcome data); NR, not reported. The Manuscript Quality Assessment Instrument (MQAI) is a 34-item scale, each item scored from 1 to 5, aimed to evaluate the quality of the research report. The Review Quality Instrument (RQI) is an 8-item scale, each item scored from 1 to 5, aimed to evaluate the quality of the peer review report

[a]Reviewers with a median score of 3 or lower on a quality scale of 1 to 5 routinely used by editors for the reviews they performed in the previous 2 years

[b]Reviewers with a median score of 4 or lower on a quality scale of 1 to 5 routinely used by editors for the reviews they performed in the previous 2 years

[c]Peer reviewers provided three peer review reports for three different manuscripts; we selected the assessment of the last manuscript

[d]We did not consider the group: "clinical reviewer + statistician + checklist reviewer", because we were interested in the effect of the use of a checklist or adding a statistician reviewer, not the effect of both

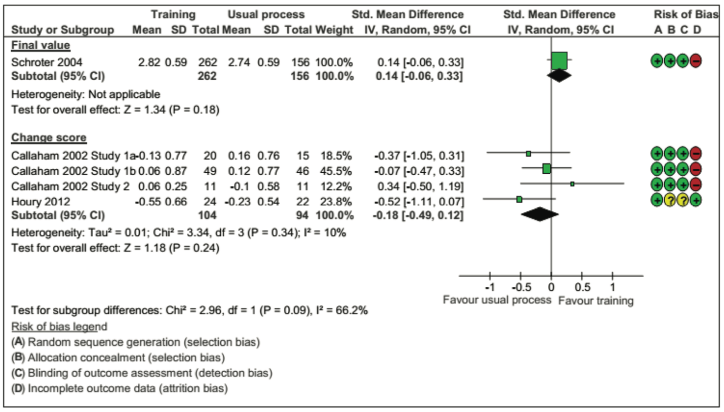



**Fig. 2** Impact of training versus usual process: standardized mean difference (SMD) of the quality of the peer review report

same comparator as for other RCTs [44]. In fact, authors compared signed peer reviews posted on the web versus signed peer reviews for authors and other peer reviewers. A total of 1702 peer reviewers, 1252 manuscripts, and 1182 manuscripts were included in the meta-analysis of the quality of peer review reports, the rejection rate, and the time peer reviewers spent on peer review, respectively. With open peer review, the quality of the peer review report increased (combined SMD, 0.14; 95 % CI, 0.05 to 0.24, $I^2 = 0$ %; Fig. 5a) and the recommendation to reject manuscripts for publication decreased, but the heterogeneity across studies was high (OR, 0.56; 95 % CI, 0.33 to 0.94; $I^2 = 67$ %; Fig. 5b). Subgroup analysis by type of open peer review (i.e., open to other peer reviewers or to authors) did not show evidence of difference. Three of the included studies evaluated the time peer reviewers spent on the peer review for each manuscript (Fig. 5c). Overall, the time peer reviewers spent on the peer review did not differ between the open peer-review and standard peer review groups (combined MD, 0.18; 95 % CI, −0.06 to 0.43, $I^2 = 50$ %). Subgroup analysis by type of open peer review (i.e., open to other peer reviewers or open to authors) did not show evidence of difference.

### *Blinded peer review*

Among six RCTs evaluating blinded peer review, three evaluated the impact on the quality of the peer review report and three the rejection rate [47–52]; 1024 and 564 manuscripts, respectively, were included in the meta-analyses. Blinded peer review did not affect the quality of the peer review report (SMD, 0.12; 95 % CI, −0.12 to 0.36, $I^2 = 68$ %) or the rate of rejection (OR, 0.77; 95 % CI, 0.39 to 1.50, $I^2 = 69$ %), but the confidence interval was large and the heterogeneity among studies substantial (Fig. 6a and b).

### *Interventions to accelerate the peer review process*

For three RCTs evaluating interventions to improve the duration of the peer review process [53–55], the interventions consisted of first asking the peer reviewers whether they agreed to review the manuscript versus directly sending the manuscript to review [53] or calling peer reviewers versus sending the manuscript [55]. Finally, the

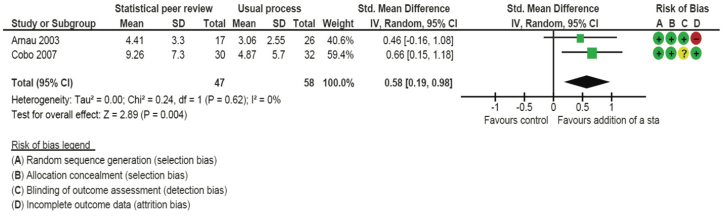



**Fig. 3** Impact of adding a statistical peer review versus usual process: standardized mean difference (SMD) of the final manuscript quality

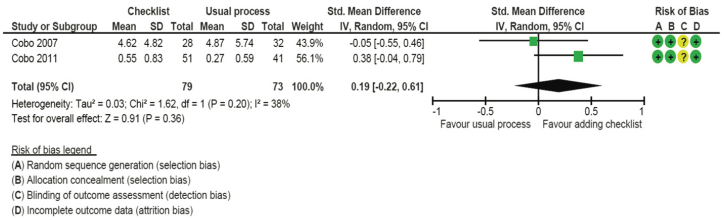



**Fig. 4** Impact of using checklist versus usual process: standardized mean difference (SMD) of the final manuscript quality

mean duration of the peer review process was decreased with an intervention based on early screening of manuscripts by editors versus the usual process [54]. We could not perform a meta-analysis because outcomes were incompletely reported.

## Discussion

Despite the essential role of peer review in the biomedical research enterprise, our systematic review identified only 22 RCTs evaluating interventions to improve the peer review quality; results of only seven RCTs were published over the past 10 years. To our knowledge, no ongoing trial is registered at trial registries. Our results provide little evidence of an effect of training of peer reviewers and the use of a checklist on the quality of the peer review report and the final manuscript. However, the amount of evidence is scarce and the methodological quality raises some concerns. Only two RCTs evaluated the impact of adding a peer reviewer with statistical expertise and showed favorable results on the quality of the final manuscript. However, these studies were performed in a single journal and need to be reproduced in other contexts. Open peer review, routinely implemented by several journals such as the *British Medical Journal* or BioMed Central journals, had a small favorable impact on the quality of the peer review report, increased the time peer reviewers spent on the review, and decreased recommendations to reject manuscripts. Blinded peer review i.e. peer reviewers are blinded to author names and affiliation did not affect the rate of rejection. The goal of blinding peer review is not to change the quality of the review but to favor objective and fair review. However, evaluating whether a review is fair and objective is a difficult or even impossible task. This explains why researchers focused on the quality of the report and the rejection rate. Nevertheless, the lack of difference observed in these outcomes should not be interpreted as a lack of usefulness of blinding. Finally, interventions to improve the speed of the process were heterogeneous.

The first International Congress on Peer Review in Biomedical Publication was held in Chicago in 1989 [56], and aimed to stimulate research on peer review. In 2007, Jefferson et al. [23] published a Cochrane systematic review on editorial peer review and concluded that a large, well-funded program of research on the effects of editorial peer review was urgently needed. Several years later, despite some initiatives such as the European COST action PEERE [57], research involving experimental design is scarce in this field, which is alarming considering the cost and central role of this process in biomedical science.

The methodological problems in studying peer review are many and complex. First, the primary functions of peer review are poorly defined. According to the editor, the role of peer reviewer varies from assessing the novelty of the study results, to assessing the scientific rigor, to assessing the clinical relevance, etc. [7]. The choice of the design is also difficult. In our review, the study designs involved the randomization of peer reviewers to assess a sample of manuscripts submitted to a journal or fabricated manuscripts developed for the RCT or the randomization of manuscripts to be assessed by a sample of peer reviewers. The selection of journals, peer reviewers and manuscripts will affect the applicability of the study results. Most RCTs were performed in a single journal and limited the generalizability of results to very specific contexts. Journal culture, standard review procedures, sub-specialties, and the experience of peer reviewers all play a role in the changed quality of reviews observed. For example, statistical peer review was studied in only one publication (*Medicina Clínica*), in three RCTs conducted by the same study team; although the results are reliable, the external validity is limited.

Bias related to the lack of blinding is also problematic when designing these RCTs. Concealing study aims and procedures from study subjects is difficult and even impossible for some of these interventions. For instance, for RCTs of participants randomized to be blinded to author identity in blinding interventions, unintentional discovery or speculation of the author's identity was reported. Similarly, due to controversial opinions of open peer review, some of the studies for open peer-review interventions were subject to bias because of a large number of potential subjects declining to participate or not completing the allocated peer review.

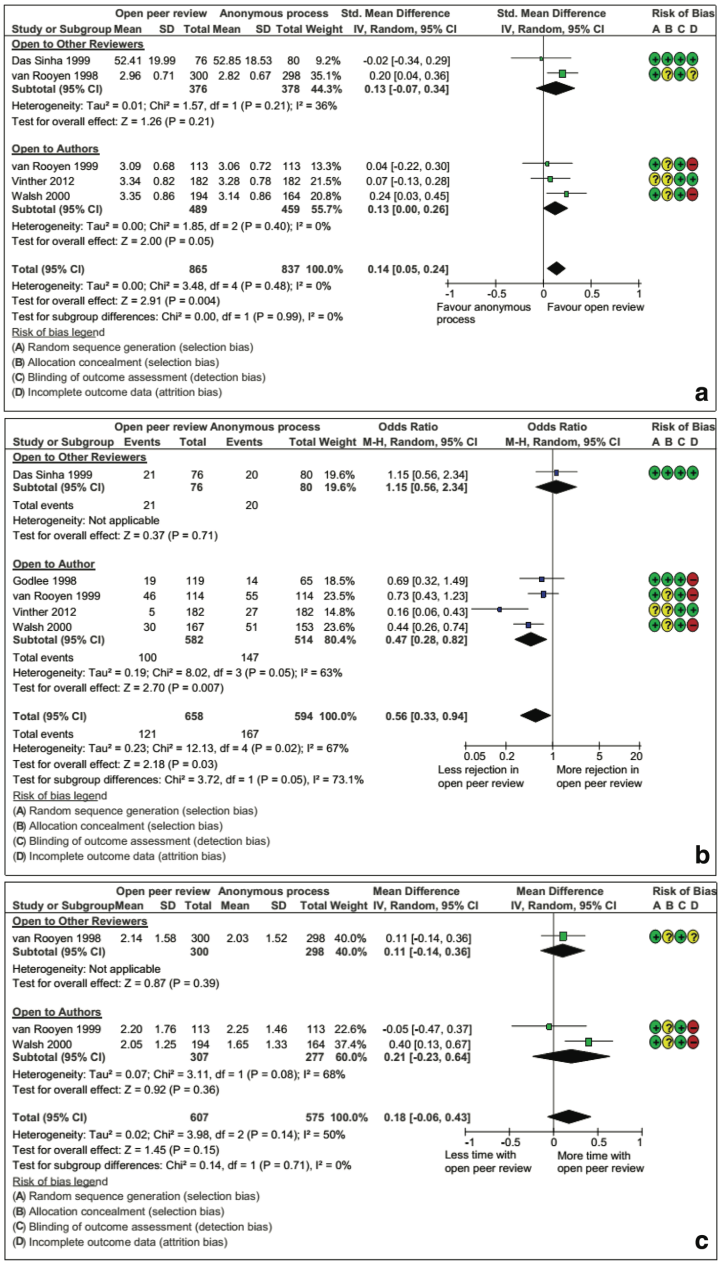



**Fig. 5** (See legend on next page.)

(See figure on previous page.)
**Fig. 5 a** Impact of the "open" review interventions versus anonymous process (anonymous to reviewers, authors or public): standardized mean difference (SMD) of the quality of the peer review report. **b** Impact of the "open" review interventions versus anonymous process (anonymous to reviewers, authors or public): odds ratio (OR) of peer reviewers' recommendation for rejection. **c** Impact of the "open" review interventions versus anonymous process (anonymous to reviewers, authors or public): standardized mean difference (SMD) of the time peer reviewers spent on the peer review

The choice of relevant outcomes is also difficult in this domain. The outcomes assessed were mainly the editor's subjective assessment or use of validated scales to assess the quality of the peer review report [35–38] or the final manuscript [39–41]. However, the choice of the criteria to consider the appropriateness of a peer review report can be debatable and can vary according to the editor in charge. For example, a study showed a discrepancy between editors and peer reviewers regarding items considered most important when assessing the report of an RCT [58]. Further, the validity of some of these outcomes (e.g., editors' routine scales) and the minimal editorial relevant difference could be questioned. Likewise, the association between recommendations to reject and the quality of the peer review is not well understood, and the proposed goal of these interventions in terms of rejection rate is not well defined. A high rejection rate could be related to the quality of the manuscript but also to the interest of the topic or to peer reviewer subjectivity. In our review, no studies evaluated the association between a recommendation of rejection and the quality of the published study. Finally, the quality of the final manuscript is not solely the results of the peer reviewers' assessment. In fact, some relevant peer reviewer suggestions might not have been acted upon, either because the authors refused or the editor did not enforce them. Of note, articles included in our systematic review evaluated the final manuscript quality after a single round of peer review. Further, for some journals, copy editing will also impact the article quality.

Our systematic review has some limitations. We selected only RCTs performed in the field of biomedical research and we cannot exclude that some interventions were developed and evaluated in other fields. We did not search some more specific bibliographic databases such as Cumulative Index to Nursing and Allied Health Literature (CINAHL). However, our search was extensive and performed according to Cochrane standards.

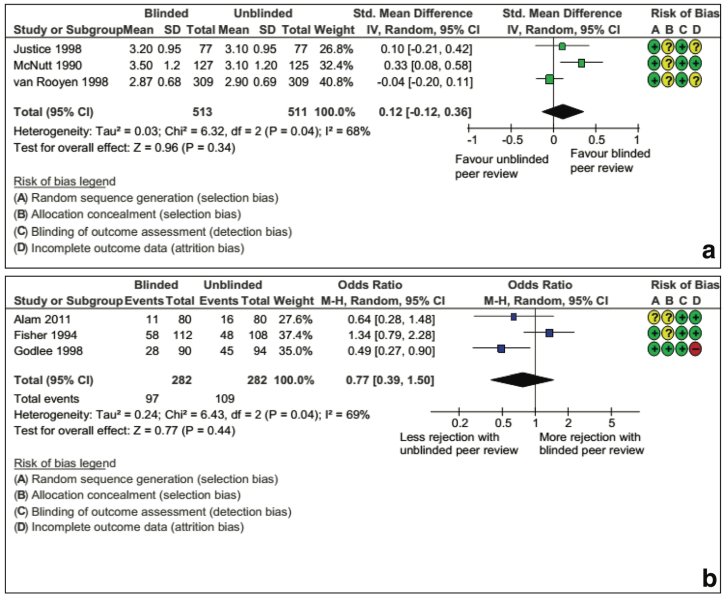


**Fig. 6 a** Impact of blinded peer review interventions versus usual process: standardized mean difference (SMD) of the quality of the peer review report. **b** Impact of blinded peer review interventions versus usual process: odds ratio (OR) of peer reviewers' recommendation for rejection

The limited number of RCTs identified, their small sample sizes, their methodological quality, and their applicability limit the interpretation of our results. Finally, reporting bias can be a concern in systematic reviews and because of the limited number of RCTs, we could not explore reporting bias. However, in this particular area, both negative and positive results are of interest to the peer-reviewed academic journal community. It is unlikely that additional RCTs meeting inclusion criteria were unreported or that results from the RCTs were selectively omitted. Still, the limitations of this body of evidence prohibit our ability to make definitive general recommendations regarding their implementation in peer review editorial processes for biomedical publications.

## Conclusions

Currently, we cannot provide conclusive recommendations on the use of interventions to improve quality of peer review from this body of evidence and its limitations. Since the publication of a previous systematic review on the topic [23], even with the inclusion of a number of more recent RCTs, the state of the evidence falls short of generating empirical support. Editorial boards and decision-makers of the publication process structure should remain aware of the uncertainty in these intervention methods and weigh the shortcomings and practical challenges before implementing them. These results highlight the urgent need to clarify the goal of the peer review process, the definition of a good quality peer review report, and the outcomes that should be used. In the longer term, these results reiterate the need for additional experimentation on this topic and exploration into the drivers of publication quality. In fact, the evidence base for recommendations needs to be strengthened. Further research is needed to identify interventions that could improve the process and assess those interventions in well-performed trials.

## Additional files

**Additional file 1: Appendix 1.** Data extraction and assessment form. (DOC 128 kb)

**Additional file 2: Appendix 2.** General characteristics of the RCTs selected. (DOC 94 kb)

**Additional file 3: Appendix 3.** Risk of bias summary: the review authors' judgments about each risk of bias item for each included study. (DOC 278 kb)

### Abbreviations

MD, mean difference; OR, odds ratio; RCT, randomized controlled trials; SMD, standardized mean difference

### Acknowledgments

We thank Erik Cobo (Universitat Politècnica Catalunya, Barcelona, Spain); Nick Black (BMJ editorial office and Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, UK); Sara Schroter (*BMJ* editorial office); Savitri Das Sinha (All India Institute of Medical Sciences, Ansari Nagar, New Delhi, India); and Michael L. Callaham (University of California, San Francisco, CA), the authors who responded to our solicitation.

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Availability of data and materials

The dataset(s) supporting the conclusions of this article not included in the appendices of this manuscript are available upon request.

### Authors' contributions

Conception and design: RB, AC, LT, PR and IB. Acquisition of data: RB and AC. Analysis: LT. Interpretation of data: RB, AC, LT, PR and IB. Drafting the article: RB, AC and IB. Revising it critically for important intellectual content: PR and LT. Final approval of the version to be published: RB, AC, LT, PR and IB.

### Competing interests

The authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no support from any organization other than the funding agency listed above for the submitted work, no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work.

### Ethics approval and consent to participate

Not required.

### Author details

[1]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA. [2]INSERM, UMR 1153, Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité – (CRESS), METHODS team, Paris, France. [3]Paris Descartes University, Paris, France. [4]Emergency department, Hôpital Lariboisière, Assistance Publique des Hôpitaux de Paris, University Diderot, Paris, France. [5]Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu Escalier A2, 1 er étage, Assistance Publique des Hôpitaux de Paris, 1 Place du Parvis Notre-Dame, Paris 75004, France.

Received: 30 April 2016 Accepted: 28 May 2016
Published online: 10 June 2016

### References

1. Kronick DA. Peer review in 18th-century scientific journalism. JAMA. 1990; 263(10):1321–2.
2. Smith R. Peer review: reform or revolution? BMJ. 1997;315(7111):759–60.
3. Rennie D. Suspended judgment. Editorial peer review: let us put it on trial. Control Clin Trials. 1992;13(6):443–5.
4. Rennie R. Editorial peer review: its development and rationale. In: Godlee F, Jefferson T, editors. Peer review in health sciences. 2nd ed. London: BMJ Books; 2003. p. 1–13.
5. Public Library of Science. Peer review—optimizing practices for online scholarly communication. In: House of Commons Science and Technology Committee, editor. Peer Review in Scientific Publications, Eighth Report of Session 2010–2012, Vol. I: Report, Together with Formal, Minutes, Oral and Written Evidence. London: The Stationery Office Limited; 2011. p. 174–8.
6. Association of American Publishers. Digital licenses replace print prices as accurate reflexion of real journal costs. 2012. http://publishers.org/sites/default/files/uploads/PSP/summer-fall_2012.pdf. Accessed 06 June 2016.
7. Jefferson T, Wager E, Davidoff F. Measuring the quality of editorial peer review. JAMA. 2002;287(21):2786–90.
8. Smith R. Peer review: a flawed process at the heart of science and journals. J R Soc Med. 2006;99(4):178–82.
9. Baxt WG, Waeckerle JF, Berlin JA, et al. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. Ann Emerg Med. 1998;32(3 Pt 1):310–7.
10. Kravitz RL, Franks P, Feldman MD, et al. Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? PLoS One. 2010;5(4):e10072.
11. Henderson M. Problems with peer review. BMJ. 2010;340:c1409.
12. Yaffe MB. Re-reviewing peer review. Sci Signal. 2009;2(85):eg11.

13. Stahel PF, Moore EE. Peer review for biomedical publications: we can improve the system. BMC Med. 2014;12(1):179.
14. Ghimire S, Kyung E, Kang W, et al. Assessment of adherence to the CONSORT statement for quality of reports on randomized controlled trial abstracts from four high-impact general medical journals. Trials. 2012;13:77.
15. Boutron I, Dutton S, Ravaud P, et al. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA. 2010;303(20):2058–64.
16. Hopewell S, Collins GS, Boutron I, et al. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. BMJ. 2014;349:g4145.
17. Turner EH, Matthew AM, Linardatos E, et al. Selective publication of antidepressant trials and its confluence on apparent efficacy. N Engl J Med. 2008;358(3):252–60.
18. Melander H, Ahlqvist-Rastad J, Meijer G, et al. Evidence b(i)ased medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ. 2003;326(7400):1171–3.
19. Lazarus C, Haneef R, Ravaud P, et al. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. BMC Med Res Methodol. 2015;15:85.
20. Jefferson T, Alderson P, Wager E, et al. Effects of editorial peer review: a systematic review. JAMA. 2002;287(21):2784–6.
21. Galipeau J, Moher D, Campbell C, et al. A systematic review highlights a knowledge gap regarding the effectiveness of health-related training programs in journalology. J Clin Epidemiol. 2015;68(3):257–65.
22. White IR, Carpenter J, Evans S, et al. Eliciting and using expert opinions about dropout bias in randomized controlled trials. Clin Trials. 2007;4(2):125–39.
23. Jefferson T, Rudin M, Brodney Folse S, et al. Editorial peer review for improving the quality of reports of biomedical studies. Cochrane Database Syst Rev. 2007;2:MR000016.
24. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: Wiley-Blackwell, 2011. Chapter 6.4.11.1. http://handbook.cochrane.org/.
25. Goodman SN, Berlin J, Fletcher SW, et al. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. Ann Intern Med. 1994;121(1):11–21.
26. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134:663–9.
27. van Rooyen S, Black N, Godlee F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. J Clin Epidemiol. 1999;52(7):625–9.
28. Callaham M, Baxt W, Waeckerle J, et al. The reliability of editors' subjective quality ratings of manuscript peer reviews. JAMA. 1998;280:229–31.
29. Black N, van Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer and a good review for a general medical journal? JAMA. 1998; 280(3):231–3.
30. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
31. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: Wiley-Blackwell, 2011. Chapter 7.6. http://handbook.cochrane.org/.
32. Borman I. DigitizeIt software v2.1. http://www.digitizeit.de/index.html. Accessed 6 June 2016.
33. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: Wiley-Blackwell, 2011. Chapter 9.5. http://handbook.cochrane.org/.
34. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: Wiley-Blackwell, 2011. Chapter 9.4.5.2. http://handbook.cochrane.org/.
35. Schroter S, Black N, Evans S, et al. Effects of training on quality of peer review: randomised controlled trial. BMJ. 2004;328(7441):673.
36. Callaham ML, Knopp RK, Gallagher EJ. Effect of written feedback by editors on quality of reviews: two randomized trials. JAMA. 2002;287(21):2781–3.
37. Callaham ML, Schriger DL. Effect of structured workshop training on subsequent performance of journal peer reviewers. Ann Emerg Med. 2002; 40(3):323–8.
38. Houry D, Green S, Callaham M. Does mentoring new peer reviewers improve review quality? A randomized trial. BMC Med Educ. 2012;12:83.
39. Arnau C, Cobo E, Ribera JM, et al. [Effect of statistical review on manuscript quality in Medicina Clinica (Barcelona): a randomized study]. Med Clin (Barc). 2003;121(18):690–4.
40. Cobo E, Selva-O'Callagham A, Ribera JM, et al. Statistical reviewers improve reporting in biomedical articles: a randomized trial. PLoS One. 2007;2(3):e332.
41. Cobo E, Cortes K, Ribera J, 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.
42. 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(5):210–3.
43. Van Rooyen S, Godlee F, Evans S, et al. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. BMJ. 1999;318:23–7.
44. Van Rooyen S, Delamothe T, Evans SJ. Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial. BMJ. 2010;341:c5729.
45. Vinther S, Nielson OH, Rosenberg J, et al. Same review quality in open versus blinded peer review in "Ugeskrift for Laeger". Dan Med. 2012;59(8):A4479.
46. Walsh E, Rooney M, Appleby L, et al. Open peer review: a randomised trial. Br J Psychiatry. 2000;176:47–51.
47. 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(3):237–40.
48. Van Rooyen S, Godlee F, Evans S, et al. Effect of blinding and unmasking on the quality of peer review. JAMA. 1998;280(3):234–7.
49. Alam M, Kim NA, Havey J, et al. Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: a randomized multi-rater study. Br J Dermatol. 2011;165:563–7.
50. Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. JAMA. 1994;272(2):143–6.
51. Justice AC, Cho MK, Winker MA, et al. Does masking author identity improve peer review quality? PEER Investigators. JAMA. 1998;280(3):240–3.
52. McNutt RA, Evans AT, Fletcher RH, et al. The effects of blinding on the quality of peer review. A randomized trial. JAMA. 1990;263(10):1371–6.
53. Pitkin RM, Burmeister LF. Identifying manuscript reviewers: randomized comparison of asking first or just sending. JAMA. 2002;287(21):2795–6.
54. Johnston SC, Lowenstein DH, Ferriero DM, et al. Early editorial manuscript screening versus obligate peer review: a randomized trial. Ann Neurol. 2007; 61(4):A10–2.
55. Neuhauser D, Koran CJ. Calling medical care reviewers first: a randomized trial. Med Care. 1989;27(6):664–6.
56. Rennie D, Knoll E, Flangrin A. The international congress on peer review in biomedical publication. JAMA. 1989;261(5):749.
57. COST European cooperation in science and technology. New Frontiers of Peer Review (PEERE). http://www.cost.eu/COST_Actions/tdp/TD1306. Accessed 6 June 2016.
58. Chauvin A, Ravaud P, Baron G, et al. 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 Med. 2015;13:158.



Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



## 4.8 Appendices de l'article

**Appendix 1. Data Extraction & Assessment Form**

| **Study ID:** | | Data extractor: □ RB □ AC |
|---|---|---|
| First Author: | Year of study publication: | |
| Title:<br><br>Publication: | | |
| Registered? □ yes □ no | Registry: | Registry #: |

**1. General Information:**

| | | | |
|---|---|---|---|
| Extracted from | □ full-text article □ abstract □ protocol □ other (specify_____) | | |
| Language | □ English □ other (specify_____) | | |
| Funding source of study | □ public □ private □ both □ none specified □ unclear | | |
| □ monocenter □ multicenter (specify # of centers_____) □ not reported<br>Countries:<br><br>Center type: □ journal □ medical school □ residency program □ other (specify_____) | | | |
| Trial design | □ 2-arm RCT □ cluster RCT □ other (specify_____) | | |
| Recruitment period | | Randomized *(total #, rate)* | |
| Study completion *(total #, rate)* | | Intervention adherence *(total #, rate)* | |

**2. Participants:**

| | **Control** | **Intervention Arm (1)** | **Intervention Arm (2)** |
|---|---|---|---|
| Age *(mean, range, SD)* | | | |
| Sex *(male:female ratio)* | | | |
| Country | | | |
| Academic rank | | | |
| Editorial board position | □ yes □ no □ unclear | □ yes □ no □ unclear | □ yes □ no □ unclear |

| | | | |
|---|---|---|---|
| Grant reviewer | □ yes □ no<br>□ unclear | □ yes □ no<br>□ unclear | □ yes □ no<br>□ unclear |
| IRB reviewer | □ yes □ no<br>□ unclear | □ yes □ no<br>□ unclear | □ yes □ no<br>□ unclear |
| Years in research *(mean, range, SD)* | | | |
| Area(s) of expertise | | | |
| Journal type & impact level *(if applicable)* | | | |

**3. Intervention(s):**

| **Control** | **Control** | **Intervention (1)** | **Intervention (2)** |
|---|---|---|---|
| Supervision | □ supervised<br>□ unsupervised<br>□ both □ unclear | □ supervised<br>□ unsupervised<br>□ both □ unclear | □ supervised<br>□ unsupervised<br>□ both □ unclear |
| Teaching style | □ passive □ hands-on<br>□ both □ unclear | □ passive □ hands-on<br>□ both □ unclear | □ passive □ hands-on<br>□ both □ unclear |
| Mode of teaching | □ face-to-face<br>□ online<br>□ print<br>□ combination<br>□ unclear | □ face-to-face<br>□ online<br>□ print<br>□ combination<br>□ unclear | □ face-to-face<br>□ online<br>□ print<br>□ combination<br>□ unclear |
| Frequency | | | |
| Duration | | | |
| Intensity *(if applicable)* | | | |
| Resources requirements | | | |
| Co-interventions | | | |
| Duration of follow-up | | | |

**3. Comparator:**

| **Control/comparison** |
|---|
| □ no intervention □ other (specify:_____) |

**4. Outcomes:**

| | | | | | |
|---|---|---|---|---|---|
| **Primary Outcome: Quality of the Review (Manuscript or peer review report)** | | | | | |
| □ RQI □ other (specify:_____) | | | | | |
| □ population-level measure □ individual-level measure | | | | | |
| Primary outcome of the study? □ yes □ no | | | | | |
| Time points measured | | | | | |
| Time points reported | | | | | |
| Unit of measurement | | | | | |
| □ Continuous □ Categorical (unordered) □ Categorical (ordered) □ Binary | | | | | |
| Mean | | Standard deviation | | Standard mean difference | |
| **Primary Outcome: Rejection rate** | | | | | |
| □ population-level measure □ individual-level measure | | | | | |
| Primary outcome of the study? □ yes □ no | | | | | |
| Time points measured | | | | | |
| Time points reported | | | | | |
| Unit of measurement | | | | | |
| □ Continuous □ Categorical (unordered) □ Categorical (ordered) □ Binary | | | | | |
| Mean | | Standard deviation | | Standard mean difference | |
| **Primary Outcome: Time spent on the peer review** | | | | | |
| □ population-level measure □ individual-level measure | | | | | |
| Primary outcome of the study? □ yes □ no | | | | | |
| Time points measured | | | | | |
| Time points reported | | | | | |
| Unit of measurement | | | | | |
| □ Continuous □ Categorical (unordered) □ Categorical (ordered) □ Binary | | | | | |
| Mean | | Standard deviation | | Standard mean difference | |
| **Other Outcomes** | | | | | |
| Specify: | | | | | |
| □ population-level measure □ individual-level measure | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Primary outcome of the study? □ yes □ no | | | | | |
| Time points measured | | | | | |
| Time points reported | | | | | |
| Unit of measurement | | | | | |
| □ Continuous □ Categorical (unordered) □ Categorical (ordered) □ Binary | | | | | |
| Mean | | Standard deviation | | Standard mean difference | |

**5. Risk of Bias**

| | |
|---|---|
| **Selection** | |
| Random sequence generation | □ high risk □ low risk □ unclear |
| Allocation concealment | □ high risk □ low risk □ unclear |
| **Performance** | |
| Who is blinded? | □ participants □ researchers □ outcome assessors □ statisticians |
| Method for blinding participants and research team | |
| Blinding | □ high risk □ low risk □ unclear |
| **Detection** | |
| Primary outcome measure | □ objective □ subjective |
| Method for blinding primary outcome assessors | |
| *Primary outcome* | □ high risk □ low risk □ unclear |
| Secondary outcome measure | □ objective □ subjective |
| Method for blinding secondary outcome assessors | |
| *Secondary outcome* | □ high risk □ low risk □ unclear |
| Other outcome measure | □ objective □ subjective |
| Method for blinding other outcome assessors | |
| *Other outcome* | □ high risk □ low risk □ unclear |
| **Attrition** | |
| Percent/number missing data in control group | |
| Reason(s) for missing data in control group | |
| Percent/number missing | |

| | |
|---|---|
| data in intervention arm 1 | |
| Reason(s) for missing data in intervention arm 1 | |
| Percent/number missing data in intervention arm 2 | |
| Reason(s) for missing data in intervention arm 2 | |
| Appropriate analysis of missing data? | □ yes □ no □ unclear |
| Attrition bias | □ high risk □ low risk □ unclear |
| **Reporting** | |
| Reported primary outcome same as protocol (or methods)? | □ yes □ no □ unclear |
| Reporting bias | □ high risk □ low risk □ unclear |

Appendix 2. General characteristics of the RCTs selected

| Author, Year | Intervention |
|---|---|
| **Training/Mentoring/Feedback** | |
| Callaham, JAMA 2002a<br><br>Study 1 | • Intervention group: written feedback by editors, i.e, peer reviewers received<br>o Other peer reviewers peer review reports and the editor's decision letter (usual process)<br>o a brief summary of the goals for a high quality peer review<br>o the editor's rating of their peer review report (feedback)<br>• Comparator group: peer reviewers received<br>o other peer reviewers peer review reports and the editor's decision letter (usual process) |
| Callaham, JAMA 2002a<br><br>Study 2 | • Intervention group: written feedback by editors, i.e, peer reviewers received<br>o other peer reviewers peer review reports and the editor's decision letter (usual process)<br>o a brief summary of the goals for a high quality peer review<br>o the editor's ratings of their peer review report<br>o the editor's ratings of the other peer reviewers' reports<br>o a copy of an excellent peer review report of another manuscript<br>• Comparator group: peer reviewers received<br>o Other peer reviewers peer review reports and the editor's decision letter |
| Callaham, 2002 Ann Emerg Med<br><br>Study 2 | • Intervention group: a 4-hour interactive workshop on peer-review run by two senior editors, consisting of presentations and discussion about peer-review<br>• Comparator group: peer reviewers did not participated in the workshop |
| Houry, 2012 BMC Med Educ | • Intervention group: standard journal practices + mentoring. Mentees were asked to discuss their peer review with their mentor by email or telephone. Mentors were asked to give feedback about how well the mentees addressed the key elements of a good peer review, what they might have done differently, and how they would rate their own peer review using the editors 1 to 5 point quality rating scale.<br>Mentors were senior peer reviewers (i.e. top 50 peer reviewers for at least 2 of the past 4 years) who agreed to be a mentor<br>• Comparator group: standard journal practices, i.e. the new peer reviewers received a packet of information materials and were encouraged to complete an online peer review training module developed by the journal |

| Author, Year | Intervention |
|---|---|
| Schroter, 2004 BMJ | • Intervention group 1: face to face training<br>- a full-day face-to-face training session on critical appraisal of clinical trials<br>- written instructions<br>- CD rom with techniques on critical appraisal of RCTs<br>• Intervention group 2: self training<br>- a self-training package consisting of the same or similar materials used in the workshop intervention.<br>• Comparator group: peer reviewers did not have any training about peer-review<br><br>We combined the two interventions groups |
| **Statistical peer review** | |
| Arnau, 2003 Med Clin (Barc) | • Intervention group: addition of a statistical peer reviewer<br>• Comparator group: usual process |

| Author, Year | Intervention |
|---|---|
| Cobo, 2007 PLoS One | 2*2 factorial design<br>• Intervention group 1: manuscripts assessed by 2 clinical peer reviewers chosen among the usual pool of peer reviewers used by the journal plus an additional statistical peer reviewer selected among the journal referee pool of 39 experts<br>• Intervention group 2: manuscripts assessed by 2 clinical peer reviewers chosen among the usual pool of peer reviewers used by the journal that received a standard letter providing the reporting guidelines relevant to the study<br>• Intervention group 3: manuscripts assessed by 2 clinical peer reviewers chosen among the usual pool of peer reviewers used by the journal that received a standard letter providing the reporting guidelines relevant to the study (i.e CONSORT, QUOROM, STARD) plus an additional statistical peer reviewer<br>• Comparator group: manuscripts assessed by 2 clinical peer reviewers chosen among the usual pool of peer reviewers used by the journal<br><br>We selected intervention group 1 and control group for this analysis |
| **Checklist** | |
| Cobo, 2007 PLoS One | *See above*<br>We selected intervention group 2 and comparator group for this analysis |
| Cobo, 2011 BMJ | • Intervention group: the usual process plus an additional peer review by a statistical peer reviewer who provided suggestions for adherence to appropriate checklists depending on the topic, such as STROBE, CONSORT, TREND, and was looking for missing items from reporting guidelines.<br>• Comparator group: the usual process that includes two peer reviewers: either two clinical peer reviewers, or one clinical peer reviewer and one statistical or epidemiological peer reviewer. |
| **Open peer review (i.e identity of peer reviewers revealed to the authors, other peer reviewers, and/or general public)** | |

| Author, Year | Intervention |
|---|---|
| Das Sinha, 1999 Natl Med J India | • Intervention group: peer reviewers were informed that their comments would be exchanged with other reviewers of the same manuscript<br><br>• Comparator group: peer reviewers received the usual cover letter along with the manuscript |
| Godlee, 1998 JAMA | • Intervention group 1: peer reviewers received a blinded manuscript (Author's details removed) and were asked to sign their peer review report<br>• Intervention group 2: peer reviewers received a blinded manuscript and were not asked to sign their report<br>• Intervention group 3: peer reviewers received a unblinded manuscript and were asked to sign their report<br>• Intervention group 4: peer reviewers received a unblinded manuscript and were not were asked to sign their report<br>• Group 5: peer reviewers were not informed of their participation in the study. They received an unblinded manuscript and were asked to sign their report. This group was not considered in the analysis as data were not available.<br>In the analysis we pooled the groups as follow<br>- Group 1 and 3: open peer reviews<br>- Group 2 and 4: anonymous<br>Group 5 was not considered in the analysis |

| Author, Year | Intervention |
|---|---|
| Van Rooyen, 1998 JAMA | • Intervention group 1: Open peer review: A pair of peer reviewers received the same manuscript. They were asked to consent to their identity being revealed to their co peer reviewers. The pair of peer reviewers was randomly allocated to a blinded (removing authors identity details from title and acknowledgement) /unblinded version of the manuscript.<br>• Intervention group 2: "the masked group". A pair of peer reviewers received the same manuscript. They were not asked to consent to their identity being revealed. The pair of peer reviewers was randomly allocated to a blinded/unblinded version of the manuscript.<br>• Uninformed group: The peer reviewers were not informed of the study<br><br>We selected intervention group 1 and 2 for this analysis. The uninformed group was not considered in the analysis |
| Van Rooyen, 1999 BMJ | • Intervention group: peer reviewers informed their identity will be revealed to authors<br><br>• Comparator group: peer reviewers remained anonymous to authors |
| Van Rooyen, 2010 BMJ | • Intervention group: peer reviewers informed that their review would be posted online<br><br>• Comparator group: peer reviewers informed that their identity would be revealed only to authors |
| Vinther, 2012 Dan Med | • Intervention group: peer reviewers were asked to have their identity revealed to authors<br><br>• Comparator group: peer reviewers remained anonymous to authors |

| Author, Year | Intervention |
|---|---|
| Walsh, 2000 Br J Psychiatry | • Intervention group: peer reviewers' identities revealed to the manuscript authors<br>• Comparator group: peer reviewers remained anonymous as per usual procedures |
| **Blinded peer review (i.e peer reviewers are blinded of the authors name and affiliation)** | |
| Alam, 2011 Br J Dermatol | For each manuscript included:<br>• Intervention group: blinded version of the manuscript (i.e authors name, affiliations and bibliography/endnotes section removed)<br>• Comparator group: unblinded version of the manuscript (i.e with authors name affiliations and bibliography/endnotes section ) |
| Fisher, 1994 JAMA | For each manuscript included:<br>• Intervention group: blinded version of the manuscript (i.e authors name removed)<br>• Comparator group: unblinded version of the manuscript (i.e with authors name) |
| Godlee, 1998 JAMA | • Intervention group 1: peer reviewers received a blinded manuscript (Author's details removed) and were asked to sign their peer review report<br>• Intervention group 2: peer reviewers received a blinded manuscript and were not asked to sign their report<br>• Intervention group 3: peer reviewers received a unblinded manuscript and were asked to sign their report<br>• Intervention group 4: peer reviewers received a unblinded manuscript and were not were asked to sign their report<br>• Group 5: peer reviewers were not informed of their participation in the study. They received an unblinded manuscript and were asked to sign their report. This group was not considered in the analysis as data were not available.<br>In the analysis we pooled the groups as follow<br>- Group 1 and 2: open peer reviews<br>- Group 3 and 4: anonymous<br>Group 5 was not considered in the analysis |
| Justice, 1998 JAMA | • Intervention group: blinded version of the manuscript (i.e authors name and institutional identity removed)<br>• Comparator group: unblinded version of the manuscript (i.e with authors name and institutional identity) |

| Author, Year | Intervention |
|---|---|
| McNutt, 1990 JAMA | For each manuscript randomized:<br>• Intervention group: blinded version of the manuscript (i.e authors name and institutional identity removed)<br>• Comparator group: unblinded version of the manuscript (i.e with authors name and institutional identity) |
| Van Rooyen, 1998 JAMA | • Intervention group: blinded version of the manuscript submitted to reviewers. (i.e authors name and institutional identity removed)<br>• Comparator group: unblinded version of the manuscript submitted to reviewers.<br>This study also assessed masked versus unmasked review (See above in open peer review intervention) |
| **Accelerate the peer review process** | |
| Johnston, 2007 Ann Neurol | • Intervention group: manuscripts were initially evaluated by editors who decided if they send it to peer reviewers.<br>• Comparator group: systematic formal external review by outside experts. |
| Neuhauser, 1989 Medical Care | • Intervention group: editors called potential reviewers before sending them a manuscript to review.<br>• Comparator group: editors send out manuscripts for review without a prior call. |
| Pitkin, 2002 JAMA | • Intervention group: the "askfirst" group. Referees received by fax information about the manuscript and asked to indicate their willingness by return fax. If they refused or failed to respond within 3working days, a substitute chosen by the editor was contacted by fax and so until a referee agreed to the review.<br>• Comparator group: the "justsend" group. Referees assigned were mailed manuscripts and asked to return their reviews by fax or e-mail within 3 weeks; if unable to comply, they were to telephone the editorial office immediately, and a substitute was chosen. |

Appendix 3. Risk of Bias Summary: the review authors' judgments about each risk of bias item for each included study

**Training/Mentoring/Feedback**

**Callaham, JAMA 2002**
**Study 1**

**Outcome:** the quality of peer review report

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | Computer-generated randomization sequence. (StatView 5.0, version 5.02, SAS) |
| *Allocation concealment (selection bias)* | **Low risk** | Reported by authors: "Therefore all allocation was by computer algorithm". |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | Reported by authors: "All editors were blinded to study purpose and reviewer participation." |
| *Incomplete outcome data (attrition bias)* | **High risk** | Very low response rate of eligible subjects for both RCTs included in this paper. "Fifty-one reviewers were eligible for entry, but 16 (10 control subjects and 6 intervention subjects) completed insufficient rate reviews during the study period, leaving 35 with sufficient data for analysis who completed 182 reviews. |

**Callaham, JAMA 2002**
**Study 2**

**Outcome:** the quality of peer review report

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | Computer-generated randomization sequence. (StatView 5.0, version 5.02, SAS) |
| *Allocation concealment (selection bias)* | **Low risk** | Reported by authors: “Therefore all allocation was by computer algorithm”. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | Reported by authors: "All editors were blinded to study purpose and reviewer participation." |
| *Incomplete outcome data (attrition bias)* | **High risk** | Very low response rate of eligible subjects for both RCTs included in this paper. Study 2: "One hundred twenty-seven reviewers were eligible; 32 (15 control subjects and 17 intervention subjects) did not complete sufficient reviews, leaving 95 with sufficient data for analysis (324 reviews)." |

**Callaham, Ann Emerg Med 2002**
**Study 2**

**Outcome:** the quality of peer review report

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | Computer-generated randomization sequence. (StatView 5.0, SAS) |
| *Allocation concealment (selection bias)* | **Low risk** | Reported by authors: “Therefore all allocation was by computer algorithm”. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | Reported by authors: "Editor's were blinded as to the identity of study participants, whether they had attended the workshop, and the purpose of both studies." |
| *Incomplete outcome data (attrition bias)* | **High risk** | High rate of follow up |

**Houry, 2012**
**Outcome:** the quality of peer review report

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "We used computer-generated randomization to assign new reviewers to either the intervention or control group." |
| *Allocation concealment (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Unclear risk** | "Over the course of the study, editors invited reviewers in their standard fashion, without the knowledge of which new reviewers were assigned to the mentorship or control groups." It is unclear whether or not editors were not blinded at the time of outcome assessment. |
| *Incomplete outcome data (attrition bias)* | **Low risk** | 4 participants excluded from the analysis in the control group |

**Schroter, 2004**

| Bias | Authors' judgment | | Support for judgment |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | | Reported by authors: "We randomised consenting reviewers into three groups: two intervention groups and a control group. We used a stratified permuted blocks randomisation method to ensure that the groups were similar in terms of factors known to influence the quality of reviews (age, current investigators in medical research projects, postgraduate training in epidemiology, postgraduate training in statistics, and editorial board members of a scientific or medical journal)." |
| *Allocation concealment (selection bias)* | **Low risk** | | Reported by authors: "Authors sent the statistician an anonymised list of study ID numbers for reviewers who agreed to take part in the study along with key the variables (age, current investigators in medical research projects, postgraduate training in epidemiology, postgraduate training in statistics, and editorial board members of a scientific or medical journal) so he could do the stratified permuted blocks randomization". |
| *Blinding of outcome assessment (detection bias)* | **Time spent** | **Low risk** | Reported by authors: Peer reviewers were aware of their participation in the trial but they were not aware of the study hypothesis. |
| | **Rejection rate** | **Low risk** | Reported by authors: Peer reviewers were aware of their participation in the trial but they were not aware of the study hypothesis. |
| | **Quality of Peer review report** | **Low risk** | This was assessed by two editors who were blinded to the identity and group allocation of the reviewers |
| *Incomplete outcome data (attrition bias)* | **Time spent** | **High risk** | "As the difference between responders and non-responders is unknown, the impact of non-response on the conclusions cannot be definitively determined." |
| | **Rejection rate** | **High risk** | |
| | **Quality of Peer review report** | **High risk** | |

**Statistical peer review**

**Arnau 2003**

**Outcome:** the final manuscript quality

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| ***Random sequence generation (selection bias)*** | **Low risk** | Computer-generated randomization sequence.<br>Translation: The randomization list, in blocks of 4, was generated computationally by an external statistician and was filed in the offices of "Clinical Medicine". The data entry and validation was performed at the Polytechnic University of Catalonia (UPC). |
| ***Allocation concealment (selection bias)*** | **Low risk** | Reported by authors: “The methodological editor previously generated the random1 to 1 allocation list with blocks of 4 (this restriction was not known by the remaining investigators)”. |
| ***Blinding of outcome assessment (detection bias)*** | **Low risk** | From English abstract: "Two evaluators, blinded to the paper's group, assessed the quality improvement in both groups, from submission to publication using a modified version of the Goodman et al. scale." Blinding of outcome assessment, and unlikely that the blinding could have been broken. |
| ***Incomplete outcome data (attrition bias)*** | **High risk** | 82 manuscripts randomized (41 in each group), but only 43 analyzed (21 in intervention group and 22 in control group). In the intervention group the reasons of exclusion were: administrative error, arrival back after the date line, no response of statistician or no sending the revision version. In the control group, the reasons of exclusion were: administrative error and arrival back after the date line. |

**Cobo, 2007**

**Outcome:** the final manuscript quality

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "Then, manuscripts were randomly allocated (by AS) using a computer program that first stratifies by study type, and second allocates to intervention groups while minimizing differences in initial quality." |
| *Allocation concealment (selection bias)* | **Low risk** | Reported by authors: " As it was obtained by a minimization algorithm, the allocated group was generated after the paper inclusion as follows:<br>1) two clinical editors (FC and JMR) included consecutively in the study all manuscripts potentially suitable for publication;<br>2) those editors assigned the standard clinical reviewers;<br>3) the methodological editor (EC) obtained the baseline data needed for the minimization algorithm;<br>4) a third clinical editor (AS) obtained the paper allocation through the minimization algorithm; and<br>5) the two crossed interventions were added to the corresponding groups:<br>(a) the methodological editor added the statistical reviewer; and<br>(b) an editorial assistant added the standard reporting guideline suggestion." |
| *Blinding of outcome assessment (detection bias)* | **Unclear risk** | "Two evaluators independtly rated the reporting quality of manuscripts at initial submission and following peer review and revision, according to the MQAI. Both knew the initial and final status but were blinded to the intervention group."<br>"On the "presence of a statistical reviewer" question, 20 (20.2%) originals were evaluated as "I don't know". The remaining 79 originals (79.8%) were inspected with a match percentage of 60.8% (95% CI from 49.1 to 71.6%)." |
| *Incomplete outcome data (attrition bias)* | **Low risk** | 129 manuscripts were randomized: 37 in the group clinical reviewers, 31 in the group clinical + statistician reviewers, 32 in the group clinical + checklist reviewer and 29 in the group clinical + statistician + checklist. 115 manuscripts analyzed: 32 in the group clinical reviewers, 30 in the group clinical + statistician reviewers, in the group 28 clinical + checklist reviewer and in the group 25 clinical + statistician + checklist. Several sensitivity analysis was performed and produced very similar conclusions. |

**Checklist**

**Cobo, 2007**

**Outcome:** the final manuscript quality

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "Then, manuscripts were randomly allocated (by AS) using a computer program that first stratifies by study type, and second allocates to intervention groups while minimizing differences in initial quality." |
| *Allocation concealment (selection bias)* | **Low risk** | Reported by authors: "As it was obtained by a minimization algorithm, the allocated group was generated after the paper inclusion as follows:<br>1) two clinical editors (FC and JMR) included consecutively in the study all manuscripts potentially suitable for publication;<br>2) those editors assigned the standard clinical reviewers;<br>3) the methodological editor (EC) obtained the baseline data needed for the minimization algorithm;<br>4) a third clinical editor (AS) obtained the paper allocation through the minimization algorithm; and<br>5) the two crossed interventions were added to the corresponding groups:<br>(a) the methodological editor added the statistical reviewer; and<br>(b) an editorial assistant added the standard reporting guideline suggestion." |
| *Blinding of outcome assessment (detection bias)* | **Unclear risk** | Reported by authors: "Two evaluators independtly rated the reporting quality of manuscripts at initial submission and following peer review and revision, according to the MQAI. Both knew the initial and final status but were blinded to the intervention group."<br>"On the use of checklist, the evaluators were able to guess the intervention group in 65.3% (95% CI from 53.5 to 76%) of the 75 over 99 (75.8%) cases analyzed." |
| *Incomplete outcome data (attrition bias)* | **Low risk** | 129 manuscripts were randomized: 37 in the group clinical reviewers, 31 in the group clinical + statistician reviewers, 32 in the group clinical + checklist reviewer and 29 in the group clinical + statistician + checklist. But 115 manuscripts analyzed: 32 in the group clinical reviewers, 30 in the group clinical + statistician reviewers, in the group 28 clinical + checklist reviewer and in the group 25 clinical + statistician + checklist. Several sensitivity analysis was performed and produced very similar conclusions. |

**Cobo, 2011**

**Outcome:** the final manuscript quality

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "Before randomisation, all manuscripts were given an ad hoc assessment by the senior statistician (EC) using a score ranging from 1 to 9, in order to give a global measure of report quality at baseline. With these scores, we were able to use a random minimisation algorithm to balance mean differences in the ad hoc score as well as differences in study type counts (that is, intervention, longitudinal, cross sectional and other type), but not to equilibrate the overall number of manuscripts in both groups. The algorithm gave probabilities from 0.5 (in the case of indifferent allocation to one or another group) to 0.8 (if both minimisation factors indicated allocation to the same group)." |
| *Allocation concealment (selection bias)* | **Low risk** | "The second editorial decision (after peer review and before randomisation) took place without committee members knowing which papers were allocated to receive the additional review (fig 1). At later editorial decisions, committee members was the additional reviews of papers in the intervention group." |
| *Blinding of outcome assessment (detection bias)* | **Unclear risk** | Reported by authors: "The three raters (recently graduates in statistics) were truly masked to the group and they didn't have access to the reviewer comments during their evaluation (first study phase). But they commented that some modifications in the new reviewed version of the paper included sentences that (they though that) they recognized either from myself or as standard RG phrases. For this reason, authors asked them if they recognized or not the group.In applying the majority rule, the joint opinion of their 3 raters guessed the intervention group in 62% of the papers (CI95% from 51% to 72%), excluding the 50% of expected guess (if random)" |
| *Incomplete outcome data (attrition bias)* | **Low risk** | 92 manuscripts randomized (41 in the control group and 51 in the control group) and 92 manuscripts analyzed |

**Open peer review**

**Das Sinha, 1999**

| Bias | Authors' judgment | | Support for judgment |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | | Reported by authors: "Randomization was done using random number charts and the sealed envelope technique." |
| *Allocation concealment (selection bias)* | **Low risk** | | Reported by authors: "Randomization was done using random number charts and the sealed envelope technique." |
| *Blinding of outcome assessment (detection bias)* | **Rejection rate** | **Low risk** | Further the aim was to evaluate if lack of blinding impact their decision |
| | **Quality of Peer review report** | **Low risk** | "The two editors were unaware of the reviewers' names, nationality, institutional affiliations, and whether or not they had been told that their comments would be exchanged." |
| *Incomplete outcome data (attrition bias)* | **Rejection rate** | **Low risk** | No mention of differences between groups in attrition reasons, nor any methods to estimate and account for missing data. Authors did not give the standard deviation of the time taken by groups. |
| | **Quality of Peer review report** | **Low risk** | |

**Godlee, 1998**

**Outcome:** The rejection rate

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "A statistician gave each reviewer a random number. These 670 reviewers were ordered and the first 420 were selected for allocation to 5 groups in random number sequence. Four of these groups were constructed, using factorial design, to investigate the effects of blinding reviewers to the authors' identities and asking them to sign their reports." |
| *Allocation concealment (selection bias)* | **Low risk** | The person who performed randomization was independent of the review. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | "Each reviewer's report was assessed independently by an editor and an epidemiologist, neither of whom was aware of the group to which the reviewer had been allocated. Where there was disagreement, the report was re-examined and consensus was reached." |
| *Incomplete outcome data (attrition bias)* | **High risk** | "Among the 90 respondents who had been blinded to the authors' identities, 23 (26%) named the authors correctly in their report. Rate ratios were little changed when those people were excluded." |

**Van Rooyen, 1998**

| **Bias** | **Authors' judgment** | | **Support for judgment** |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | | "The randomization process was undertaken by a researcher who was independent from the editorial decision-making process using a computerized minimization program with a random component." |
| *Allocation concealment (selection bias)* | **Unclear risk** | | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Time spent** | **Low risk** | Reported by authors: "Peer reviewers were aware of their participation in the trial but they were not aware of the study hypothesis". |
| | **Quality of Peer review report** | **Low risk** | "At no stage were editors or authors aware of the group to which a manuscript has been allocated." |
| *Incomplete outcome data (attrition bias)* | **Time spent** | **Unclear risk** | "Of the 527 manuscripts entered into the study, 487 (89%) were successfully randomized and followed up". |
| | **Quality of Peer review report** | **Unclear risk** | |

**Van Rooyen, 1999**

| **Bias** | **Authors' judgment** | | **Support for judgment** |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | | "Randomisation was carried out by a researcher using a computerised randomisation program." |
| *Allocation concealment (selection bias)* | **Unclear risk** | | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Time spent** | **Low risk** | Reported by authors: "Peer reviewers were aware of their participation in the trial but they were not aware of the study hypothesis". |
| | **Rejection rate** | **Low risk** | Further the aim was to evaluate if lack of blinding impact their decision |
| | **Quality of Peer review report** | **Low risk** | "Editors did not know which of the reviewers had consented to be identified to the author." |
| *Incomplete outcome data (attrition bias)* | **Time spent** | **High risk** | Of the 250 reviewers initially invited to participate, 73 declined—29 (23%) in the group randomised to remain anonymous (anonymous reviewers) and 44 (35%) in the group randomised to be asked for consent to be identified (identified reviewers). In 11 of the 125 randomised papers both reviewers declined, in 18 only the anonymous reviewer declined, and in 33 only the identified reviewer declined. Thus, the difference between identified reviewers and anonymous reviewers in declining to review was 12% (35%v23%, 95% confidence interval 0.2% to 24%), which is marginally significant (McNemar's $\chi 2$= 3.84, P = 0.0499). |
| | **Rejection rate** | **High risk** | |
| | **Quality of Peer review report** | **High risk** | |

**Van Rooyen, 2010**

| Bias | Authors' judgment | | Support for judgment |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Unclear risk** | | No information gave by authors about the random sequence generation. |
| *Allocation concealment (selection bias)* | **Unclear risk** | | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Time spent** | **Low risk** | Peer reviewers did not seem to be aware of the study hypothesis. |
| | **Rejection rate** | **Low risk** | Further the aim was to evaluate if lack of blinding impact their decision |
| | **Quality of Peer review report** | **Low risk** | "Editors were blinded throughout to the group allocation." |
| *Incomplete outcome data (attrition bias)* | **Time spent** | **Low Risk** | 225 of 265 (85%) randomized manuscripts were analyzed in the intervention group and 246 of 283 (87%) randomized manuscripts were analyzed in the control group. |
| | **Rejection rate** | **Low risk** | |
| | **Quality of Peer review report** | **Low risk** | |

**Vinther, 2012**

| Bias | Authors' judgment | | Support for judgment |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Unclear risk** | | Insufficient information about the sequence generation process to permit judgment of 'Yes' or 'No'. |
| *Allocation concealment (selection bias)* | **Unclear risk** | | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Rejection rate** | **Low risk** | Further the aim was to evaluate if lack of blinding impact their decision |
| | **Quality of Peer review report** | **Low risk** | "Review quality was assessed by two editors, neither of whom was aware of the group to which a reviewer had been allocated." |
| *Incomplete outcome data (attrition bias)* | **Rejection rate** | **Low risk** | "For eight manuscripts, it was not possible to obtain complete data (an open and a blinded review) despite that several reminders were sent to reviewers; these eight manuscripts (corresponding to 1 reviews) were excluded from the analysis." Missing data represented less 5% in intervention and control group. |
| | **Quality of Peer review report** | **Low risk** | |

**Walsh, 2000**

| Bias | Authors' judgment | | Support for judgment |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | | "Simple randomization was performed using computer-generated random numbers." |
| *Allocation concealment (selection bias)* | **Unclear risk** | | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Time spent** | **Low risk** | Peer reviewers did not seem to be aware of the study hypothesis |
| | **Rejection rate** | **Low risk** | Further the aim was to evaluate if lack of blinding impact their decision |
| | **Quality of Peer review report** | **Low risk** | "Throughout the trial period, reviews returned by those who had declined to reveal their names to authors were rated on quality. These reviews were rated blind, interspersed with the other reviews, by the Trainee Editors." "...use of blind ratings and good inter-rater reliability as measured by weighted k statistics." |
| *Incomplete outcome data (attrition bias)* | **Time spent** | **High risk** | "It was thought that the 'decliners' group may differ systematically from the participating group in terms of review quality. An attempt was made to estimate any differences." 57 refused to participate because they did not want to be in the "open" review group. Otherwise, those who were not followed up (did not complete the review) was roughly the same in the control and "open" review groups (12% and 13% respectively). |
| | **Rejection rate** | **High risk** | |
| | **Quality of Peer review report** | **High risk** | |

**Blinded peer review**

**Alam, 2011**

**Outcome:** The rejection rate

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Unclear risk** | Insufficient information about the sequence generation process to permit judgment of 'Yes' or 'No'. |
| *Allocation concealment (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | Further the aim was to evaluate if lack of blinding impact their decision |
| *Incomplete outcome data (attrition bias)* | **Low risk** | "In a total of six cases, reviewers failed to complete their reviews within 1 month of the date of assignment; a different reviewer was then randomly assigned to replace each of these. Replacement reviewers were substituted for three blinded and three unblinded reviewers, and in no instance were multiple replacement reviewers required for a given manuscript." |

**Fisher, 1994**

**Outcome:** The rejection rate

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "Which reviewers were blinded was determined by use of a computer-generated random-numbers table." |
| *Allocation concealment (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | Further the aim was to evaluate if lack of blinding impact their decision |
| *Incomplete outcome data (attrition bias)* | **Low risk** | "Two blinded and six nonblinded reviewers did not return their reviews."<br>Concerning the rejection rate, the worth case study showed an impact of missing data. |

**Godlee, 1998**

**Outcome:** The rejection rate

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "A statistician gave each reviewer a random number. These 670 reviewers were ordered and the first 420 were selected for allocation to 5 groups in random number sequence. Four of these groups were constructed, using factorial design, to investigate the effects of blinding reviewers to the authors' identities and asking them to sign their reports." |
| *Allocation concealment (selection bias)* | **Low risk** | The person who performed randomization was independent of the review. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | "Each reviewer's report was assessed independently by an editor and an epidemiologist, neither of whom was aware of the group to which the reviewer had been allocated. Where there was disagreement, the report was re-examined and consensus was reached." |
| *Incomplete outcome data (attrition bias)* | **High risk** | "Among the 90 respondents who had been blinded to the authors' identities, 23 (26%) named the authors correctly in their report. Rate ratios were little changed when those people were excluded." |

**Justice, 1998**

**Outcome:** the quality of peer review report

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "Randomization was performed using random number tables." |
| *Allocation concealment (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | "The manuscript editor and the corresponding author rated the quality of each review, unaware of the group to which the review had been assigned." |
| *Incomplete outcome data (attrition bias)* | **Unclear risk** | "Of those randomized to the intervention, 77 (84%) of 92 had sufficient data to compare a review quality based on the editor's judgment." No mention of the missing data in the control group. |

**McNutt, 1990**

**Outcome:** the quality of peer review report

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | Sufficient information about the sequence generation process to permit judgment of 'Yes' or 'No' |
| *Allocation concealment (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | "After reviews were returned, the manuscript and the reviews were given to an editor who was blinded to the authors of the manuscript and their institutions and to the identity of the reviewers. The editor then graded the quality of both the manuscript and its reviews." |
| *Incomplete outcome data (attrition bias)* | **Low risk** | 131 manuscripts randomized, 127 in the intervention group and 125 in control group were analyzed. Missing data represented less than 10% in intervention and control group |

**Van Rooyen, 1998**

| **Bias** | **Authors' judgment** | | **Support for judgment** |
|---|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | | "The randomization process was undertaken by a researcher who was independent from the editorial decision-making process using a computerized minimization program with a random component." |
| *Allocation concealment (selection bias)* | **Unclear risk** | | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Time spent** | **Low risk** | Reported by authors: "Peer reviewers were aware of their participation in the trial but they were not aware of the study hypothesis". |
| | **Quality of Peer review report** | **Low risk** | "At no stage were editors or authors aware of the group to which a manuscript has been allocated." |
| *Incomplete outcome data (attrition bias)* | **Time spent** | **Unclear Risk** | "Of the 527 manuscripts entered into the study, 487 (89%) were successfully randomized and followed up". |
| | **Quality of Peer review report** | **Unclear risk** | |

**Accelerate the peer review**

**Johnston, 2007**

**Outcome:** accelerate the peer review process

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Allocation concealment (selection bias)* | **Low risk** | "Without knowledge of randomization group, one of our six editors made a decision about whether a manuscript should be sent out for review or rejected without further consideration". |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | The outcome was the delay of the peer review process, so the blinding was not necessary for the assessment. |
| *Incomplete outcome data (attrition bias)* | **Low risk** | All manuscripts included were analyzed. |

**Neuhauser, 1989**

**Outcome:** accelerate the peer review process

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Allocation concealment (selection bias)* | **Unclear risk** | Insufficient information to permit judgment of 'Yes' or 'No'. The method of concealment is not described or not described in sufficient detail to allow a definite judgment. |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | The outcome was the delay of the peer review process, so the blinding was not necessary for the assessment. |
| *Incomplete outcome data (attrition bias)* | **Low risk** | "Ninety-five percent of called manuscript were reviewed compared with 92% of noncalled manuscripts, an insignificant difference." |

**Pitkin, 2002**

**Outcome:** accelerate the peer review process

| Bias | Authors' judgment | Support for judgment |
|---|---|---|
| *Random sequence generation (selection bias)* | **Low risk** | "Using a random-number generator, an editorial assistant assigned one referee to justsend and the other to askfirst". |
| *Allocation concealment (selection bias)* | **Low risk** | "Using a random-number generator, an editorial assistant assigned one referee to justsend and the other to askfirst". |
| *Blinding of outcome assessment (detection bias)* | **Low risk** | The outcome was the delay of the peer review process, so the blinding was not necessary for the assessment. |
| *Incomplete outcome data (attrition bias)* | **Low risk** | All manuscripts included were analyzed. |

Appenidx 4. The PRISMA checklist

| Section/topic | # | Checklist item | Reported on page # |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review, meta-analysis, or both. | 1 |
| **ABSTRACT** | | | |
| Structured summary | 2 | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. | 3 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of what is already known. | 5 |
| Objectives | 4 | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS). | 5 |
| **METHODS** | | | |
| Protocol and registration | 5 | Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. | - |
| Eligibility criteria | 6 | Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. | 6 |
| Information sources | 7 | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. | 6 |
| Search | 8 | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. | 6 |
| Study selection | 9 | State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis). | 6 |
| Data collection process | 10 | Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators. | 9 |
| Data items | 11 | List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made. | 9/10 |
| Risk of bias in individual studies | 12 | Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. | 8 |

| | | | |
|---|---|---|---|
| Summary measures | 13 | State the principal summary measures (e.g., risk ratio, difference in means). | 9/10 |
| Synthesis of results | 14 | Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$) for each meta-analysis. | 9/10 |

| Section/topic | # | Checklist item | Reported on page # |
|---|---|---|---|
| Risk of bias across studies | 15 | Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies). | 9 |
| Additional analyses | 16 | Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. | 10 |
| **RESULTS** | | | |
| Study selection | 17 | Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram. | 11 |
| Study characteristics | 18 | For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations. | 11-15 |
| Risk of bias within studies | 19 | Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). | Figures |
| Results of individual studies | 20 | For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. | 11-15 |
| Synthesis of results | 21 | Present results of each meta-analysis done, including confidence intervals and measures of consistency. | 11-15 |
| Risk of bias across studies | 22 | Present results of any assessment of risk of bias across studies (see Item 15). | 11-15 |
| Additional analysis | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). | - |
| **DISCUSSION** | | | |
| Summary of evidence | 24 | Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers). | 16 |
| Limitations | 25 | Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias). | 18 |
| Conclusions | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research. | 19 |
| **FUNDING** | | | |

| | | | |
|---|---|---|---|
| Funding | 27 | Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. | 20 |

*From:* Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

# 5. COBPeer : un outil informatisé permettant de standardiser les critiques des reviewers sur la qualité de la présentation des articles.

## 5.1 Justification de l'étude

Les publications devraient décrire de façon transparente, complète et exacte les méthodes utilisées ainsi que les résultats. En effet, de nombreux articles omettent des informations importantes nécessaires pour évaluer leur robustesse méthodologique et leur pertinence clinique. Sans un rapport clair et complet, il n'est pas possible d'identifier les failles ou les biais de ces articles, de reproduire les interventions ou d'utiliser les résultats dans des revues systématiques ou des méta-analyses (122). Pour favoriser la transparence, des recommandations internationales, comme le CONSORT Statement, ont été développées. La plupart des éditeurs exige que les auteurs rapportent leurs articles en respectant ces recommandations. Cependant, plusieurs travaux ayant évalué la qualité de la présentation des études publiées ont démontré que les informations essentielles pour évaluer la qualité méthodologique des études et utiliser les résultats sont manquantes (123). En effet, malgré une amélioration de l'adhérence aux recommandations, celle-ci reste sous optimale, et ce malgré l'intervention des peer-reviewers (124) (125).

Par ailleurs, une présentation sélective des résultats et une modification des critères de jugement pour mettre en avant les critères de jugement statistiquement significatifs est un phénomène fréquent dans les publications (i.e., un écart entre le critère de jugement pré-spécifié dans le registre d'essai ou le protocole et celui rapporté dans le manuscrit) (126). Pour éviter ce type de biais (127), tous les critères de jugement pré-spécifiés doivent être signalés dans le manuscrit, et les auteurs doivent signaler toute modification réalisée après le début de l'étude, et ce, avec les raisons ((128).

Nos travaux ont montré que les reviewers ne considèrent pas que l'évaluation de la qualité de la présentation des publications, le respect des recommandations de présentation comme le

CONSORT Statement et la comparaison des critères de jugement rapportés dans l'article et dans les registres soient des tâches importantes (44).

Notre hypothèse est que ces tâches n'exigent pas un haut niveau d'expertise et pourraient être exécutées par des chercheurs au début de leur carrière. Ceux-ci pourraient utiliser un outil en ligne simple basé, sur la checklist du CONSORT 2010 (129).

## 5.2 Le CONSORT: *Consolidated Standards of Reporting Trials*

Les premières ébauches du CONSORT datent de 1993, lorsque 30 experts composés d'éditeurs de revues médicales, de spécialistes des essais cliniques, d'épidémiologistes et de méthodologistes se sont réunis à Ottawa, au Canada, dans le but d'élaborer une nouvelle échelle pour évaluer la qualité des essais contrôlés randomisés. Au cours des discussions préliminaires, les participants ont estimé que de nombreux items proposés n'étaient pas pertinents parce qu'ils n'étaient pas régulièrement rapportés par les auteurs. De plus en plus d'études démontraient que la qualité des rapports d'essais contrôlés randomisés était loin d'être optimale (130) (131). Par conséquent, un accord unanime a guidé le reste de l'atelier pour se concentrer sur les moyens simples pour améliorer la manière dont sont rapportés (i.e.; *reporting*) les éléments clés par les auteurs d'essais contrôlés randomisés.

Les participants ont décidé d'inclure dans la mesure du possible, dans la checklist des items pour lesquels il y avait des preuves ou sur le fait que l'absence de cette information pourrait conduire à des interprétations biaisées de l'intervention étudiée. En 1994, les auteurs se sont accordés sur une première liste : le SORT pour *The Standards of Reporting Trials Group* (132). Cette liste comprenait 32 items.

Simultanément, et de manière indépendante, un autre groupe d'experts, *Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature*, s'est réuni

aux États-Unis, pour travailler sur la même thématique. Ce groupe a également publié une proposition qui comprenait une liste de points à recommander aux auteurs lors de la rédaction de manuscrits des essais contrôlés randomisés (133).

À la suggestion de Drummond Rennie, rédacteur en chef adjoint du *Journal of American Medicine Association*, des représentants des deux groupes se sont rencontrés en 1996, à Chicago, aux États-Unis. L'objectif de ce groupe de travail était de fusionner les deux listes pour n'en faire qu'une seule qui soit cohérente et fondée sur des preuves. En effet, les experts ont estimé qu'une seule recommandation aurait plus de chances de séduire les revues et donc d'améliorer sa diffusion. La réunion a abouti à la déclaration du CONSORT pour *Consolidated Standards of Reporting Trials*, publiée pour la première fois en 1996 (134).

D'autres réunions du Groupe en 1999 et en 2000 ont conduit à la publication de la révision CONSORT 2001. Suite à une réunion en janvier 2007, une nouvelle révision a été développée et le *CONSORT statement 2010* a été publiée le 24 mars 2010.

L'intention initiale de la création de la checklist du CONSORT était de ne garder que les éléments jugés fondamentaux pour le *reporting* d'un essai contrôlé randomisé. Certains éléments, qui n'étaient pas considérés comme essentiels, mais qui pouvaient être hautement souhaitables lors de la soumission d'un essai contrôlé randomisé, n'avaient pas été initialement introduits dans le CONSORT. Ces éléments étaient : l'approbation d'un comité d'éthique institutionnel, les sources de financement pour l'essai et le numéro de registre d'essai (i.e., le numéro ISRCTN pour *International Standard Randomized Controlled Trial Number*) (135). Ces éléments ont été introduits lors de la révision du CONSORT et publiée en 2001. Pour mémoire, cette révision a été publiée simultanément dans trois prestigieuses revues médicales internationales (136).

Depuis la révision en 2001, il avait été mis en évidence de nouvelles préoccupations concernant des défauts dans le rapport d'essais contrôlés randomisés, tels que par exemple le rapport sélectif

des résultats (137). De ce fait, 31 membres du groupe CONSORT se sont réunis à Montebello, au Canada, en janvier 2007 pour mettre à jour le CONSORT de 2001. Comme dans les versions antérieures du CONSORT, seuls les éléments jugés fondamentaux pour la rédaction d'un essai contrôlé randomisé ont été conservés. Pour favoriser l'adhésion des auteurs et des éditeurs, les experts ont rédigé un article d'explication (i.e., *Explanation and elaboration*). Le 24 mars 2010, huit revues publiaient simultanément la déclaration CONSORT 2010.

De nombreux auteurs avaient mis en évidence la nécessité de promouvoir et de disséminer le CONSORT pour améliorer la qualité des essais randomisés publiés (138). Cette volonté de favoriser la transparence de la recherche pour lutter contre le gaspillage(139)de la recherche (136) fut soutenue par les éditeurs (140).

De nombreuses études ont mis en évidence une amélioration de la qualité de la description des essais contrôlés randomisés grâce à l'utilisation de la checklist CONSORT (141). Mais cette amélioration qualitative des essais reste faible et il devient donc urgent de proposer des interventions originales pour favoriser la dissémination d'une recherche de qualité (96).

## 5.3 Objectifs

Les objectifs de l'étude étaient :

1) Elaborer un outil d'aide à l'évaluation de la qualité de la présentation des essais randomisés et de la présentation sélective des résultats. Cet outil est dédiés aux peer-reviewers junior (i.e., les jeunes chercheurs, les étudiants en master, les étudiants en thèse d'Université, les

cliniciens impliqués dans la recherche clinique et qui n'ont jamais été peer-reviewers). Cet outil s'appuie sur : a) la checklist CONSORT 2010, b) le processus COMPare utilisé pour identifier les changements de critères de jugement principaux dans les essais publiés.

2) Comparer la performance des peer-reviewers junior utilisant l'outil avec la performance des peer-reviewers initialement assignés à un manuscrit, dans l'identification des items du CONSORT mal rapportés par les auteurs ainsi que des changements du critère de jugement principal.

## 5.4 Méthodes

Nous avons réalisé une étude transversale diagnostique comparant la performance des peer-reviewers junior utilisant l'outil, avec la performance des peer-reviewers initialement affectés à un manuscrit, dans l'identification des items du CONSORT mal rapportés par les auteurs et des changements du critère de jugement principal dans la première version soumise du manuscrit.

### 5.4.1 Développement de l'outil COBPeer

L'outil COBPeer, « Consort-Based Peer-Review Tool »(Figure 4), donne des conseils aux peer-reviewer juniors quand ils évaluent : 1) l'exhaustivité des items rapportés par les auteurs selon la checklist CONSORT rapportés par les auteurs, 2) le changement de critère de jugement principal. Avant d'utiliser l'outil, les peer-reviewers juniors bénéficieront d'un entrainement consistant en l'évaluation d'extrait de manuscrit pour chaque item CONSORT sélectionné ainsi que sur l'enregistrement des essais. Pour chaque exercice, les participants auront un retour selon leurs réponses données.

"Outcomes

After one-year anthropometric data, information on smoking behavior, changes in medication and medical history, fasting blood glucose, lipids and creatinine were collected. The baseline questionnaire was repeated and treatment plans were analyzed on the number and duration of visits and use of self-monitoring. For the calculation of the SCORE risk assessment both at baseline and after one year, the age at baseline was used, entailing slight underestimation of the risk after one year."

| Outcomes | Item 6a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed | | |
|---|---|---|---|
| | | Yes | No |
| | • Was the primary outcome clearly identified ? | ○ | ○ |

| Outcomes | Item 6a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed | | | |
|---|---|---|---|---|
| | | Yes | No | NA |
| | Did the author report for primary outcome: (answer yes if it is true for all primary outcomes) | | | |
| | • The variable of interest (e.g., pain, depression, all-cause mortality) | ○ | ○ | |
| | • How the outcome was assessed (e.g., VAS, Beck Depression Inventory Scale) | ○ | ○ | |
| | • The format of the primary outcome data (i.e. the analysis metric) (e.g., change from baseline, final value, time to event) | ○ | ○ | |
| | • The summary measure for each study group (e.g., mean, proportion with score > 2) | ○ | ○ | |
| | • The time frame to measure primary outcome (e.g., 3 months) **NA if survival analysis* | ○ | ○ | ○ |
| | • Who assessed the outcome (e.g., the patient, doctor, nurse, caregiver, other) | ○ | ○ | |

**Figure 4. Exemple d'exercice proposé aux participants**

### 5.4.2 Intégralité du rapport des items pour les essais contrôlés randomisés

Nous nous sommes concentrés sur les items les moins souvent rapportés par les auteurs (i.e., rapportés moins de 50% selon l'étude de Hopewel et al.(36) et nécessaire à la réalisation d'une méta-analyse. Les éléments sélectionnés sont :

- Cinq items dans la section méthode : critère de jugement (item 6a), randomisation (item 8a), mécanisme d'allocation (item 9), l'aveugle (item 11a/11b)
- Quatre items dans la section résultats : flux des participants (item 13a/13b), résultats et estimation (item 17a), effets secondaires (item 19)
- Un item dédié à l'enregistrement (item 23)

## 5.5 Interventions (Figure 5)

Nous avons identifié un échantillon de 120 articles rapportant un essai randomisé publié dans BioMed Central qui donne accès au manuscrit soumis et aux commentaires de reviewers.
Les manuscrits soumis aux journaux vont être évalués de trois manières : 1) le gold standard ; 2) les peer-reviewers junior ; 3) les peer-reviewers dans le cadre du processus usuel

Le gold standard correspond à des paires de *systematic reviewers* qui vont indépendamment extraire les données en utilisant un questionnaire standardisé. Pour l'évaluation des items retenus, le gold standard va réaliser une évaluation des risques de biais et de la description de la prise en compte des effets secondaires par les auteurs dans la première version de leur manuscrit soumise. Chaque item sera évalué en : correctement rapporté/ incorrectement rapporté. Le gold standard n'aura pas accès à l'outil COBPeer (Figure 6).

Les peer-reviewers juniors vont évaluer les mêmes items dans les mêmes manuscrits en utilisant l'outil COBPeer après avoir suivi un entrainement standardisé.

Le processus de peer-review classique : Deux chercheurs vont évaluer les rapports des peer-reviewers ayant évalué la première version du manuscrit soumis, pour déterminer si les peer-reviewers ont, oui ou non, mis en évidence des manquements de la part des auteurs dans la manière dont sont rapportés les items.

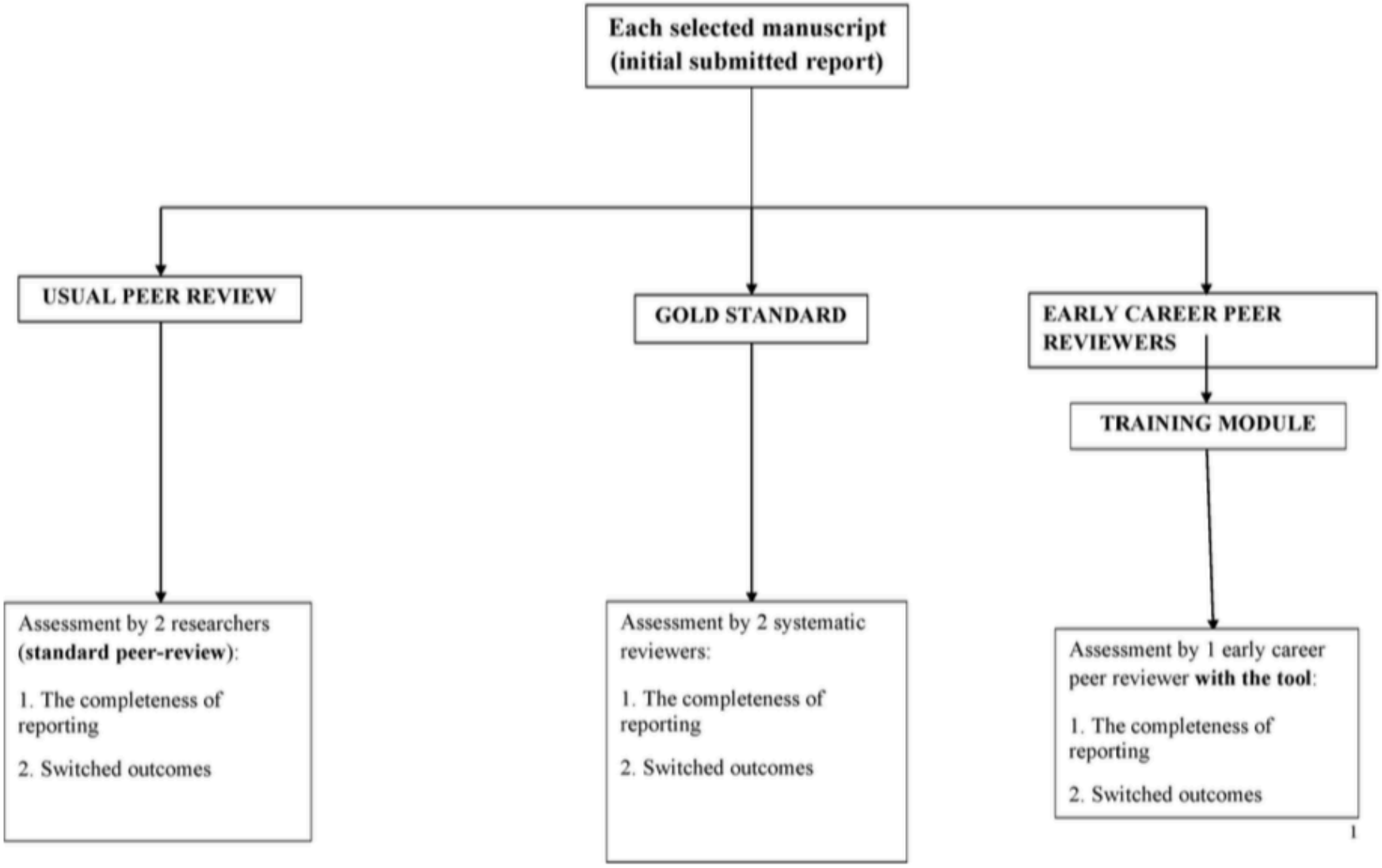



**Figure 5. Le design de l'étude**

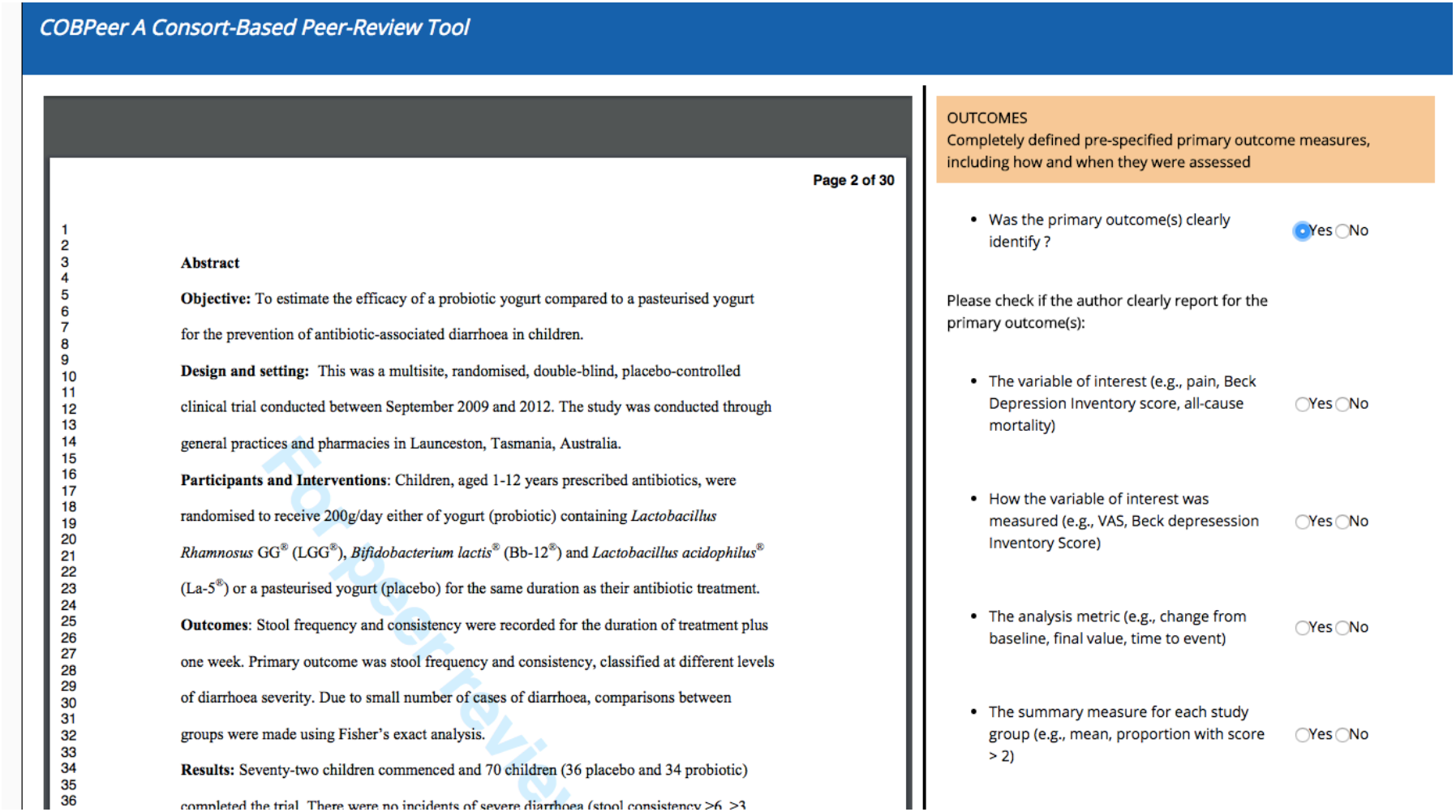



**Figure 6. Aperçu de l'outil COBPeer pour l'évaluation des essais contrôlés randomisés**

## 5.6 Les critères de jugement

Le critère de jugement principal était le nombre moyen d'items correctement classés par manuscrit.

Les critères de jugements secondaires étaient le nombre moyen d'items du CONSORT correctement classé par manuscrit et la sensitivité, spécificité ainsi que les rapports de vraisemblance pour détecter les items et les changements de critères de jugement.

## 5.7 Résultats

Le nombre moyen d'items correctement classés par manuscrit était de 6,18 (écart-type : 1,53) pour les jeunes chercheurs versus 4,74 (écart-type : 2,01) pour les peer-reviewers ayant initialement évalué le manuscrit ($p<0.01$).

Pour les critères de jugement secondaires, tout d'abord, le nombre d'items correctement classés parmi les 10 items CONSORT sélectionnés étaient de 5,61 (écart-type :1,43) pour les jeunes chercheurs versus 4,34 (écart-type : 1,81) pour les peer-reviewers ayant initialement évalué le manuscrit ($p<0.01$). De plus, en dehors des items se rapportant à l'aveugle, notre outil permet aux peer-reviewers juniors d'avoir une meilleure sensibilité dans la détection des items mal rapportés par les auteurs (Tableau 1).

**Table 1. Les résultats de l'étude COBPeer**

| | Item 6a (Outcomes) | | Item 8a (Randomization / Sequence generation) | | Item 9 (Allocation concealment mechanism) | | Items 11a/11b (Blinding) | | Items 13a/13b (Participants flow) | | Item 17a (Outcomes and estimation) | | Item 19 (Harms) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** |
| **Sensitivity** | 27.1 | 94.9 | 33.3 | 74.4 | 7.9 | 82.5 | 19.6 | 56.9 | 17.9 | 92.3 | 33.3 | 93.8 | 9.9 | 97.2 |
| **Specificity** | 57.4 | 42.6 | 76.5 | 85.2 | 87.7 | 71.9 | 88.4 | 65.2 | 79.0 | 22.2 | 65.3 | 56.9 | 75.5 | 10.2 |
| **Positive likelihood ratio** | 38.1 | 61.5 | 40.6 | 70.7 | 41.7 | 76.5 | 55.6 | 54.7 | 29.2 | 36.4 | 39.0 | 59.2 | 36.8 | 61.1 |
| **Negative likelihood ratio** | 44.9 | 89.7 | 70.5 | 87.3 | 46.3 | 78.9 | 59.8 | 67.2 | 66.7 | 85.7 | 59.5 | 93.2 | 36.6 | 71.4 |

| | Item 23 (Registration) | | Switched outcomes | | TOTAL | |
|---|---|---|---|---|---|---|
| | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** | **Usual peer review** | **Early career peer reviewer** |
| **Sensitivity** | 10.5 | 84.2 | 5.4 | 71.4 | 17.3 | 83.4 |
| **Specificity** | 88.1 | 95.1 | 95.7 | 42.1 | 78.1 | 54.9 |
| **Positive likelihood ratio** | 14.3 | 76.2 | 66.7 | 57.7 | 39.6 | 58.9 |
| **Negative likelihood ratio** | 84.0 | 97.0 | 38.6 | 57.1 | 53.1 | 81.0 |

## 5.8 Discussion et limites

Nos résultats ont mis en évidence que l'outil a amélioré le nombre d'items correctement classés par manuscrit par les peer-reviewers juniors. Cette constatation suggère que l'utilisation d'une checklist de type CONSORT se focalisant sur certains items précis pourrait améliorer la qualité du processus de peer-review, et par extension, la qualité du manuscrit publié.

Notre étude présente des limites. Premièrement, nous ne pouvons pas savoir si les chercheurs en début de carrière qui ont participé à l'étude sont représentatifs de tous les chercheurs en début de carrière. De plus, nous pourrions avoir un biais de recrutement en raison du temps requis pour terminer l'étude. Deuxièmement, l'outil en ligne n'a examiné que dix items du CONSORT 2010 et la modification du critère de jugement principal. Nous n'avons pas examiné l'impact de l'évaluation de l'ensemble des items du CONSORT et de toutes ses extensions. Troisièmement, les manuscrits sélectionnés pouvaient ne pas être représentatifs de la qualité des rapport soumis aux évaluateurs. Mais, la sélection randomisée parmi un grand panel de revue biomédicale pourrait garantir l'hétérogénéité dans la qualité du manuscrit.

## 5.9 Article

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



BMJ Open

# A protocol of a cross-sectional study evaluating an online tool for early career peer reviewers assessing reports of randomised controlled trials

Anthony Chauvin,[1,2,3] David Moher,[4,5] Doug Altman,[6] David L Schriger,[7] Sabina Alam,[8] Sally Hopewell,[9] Daniel R Shanahan,[10] Alessandro Recchioni,[10] Philippe Ravaud,[1,2] Isabelle Boutron[1,2,11]

**To cite:** Chauvin A, Moher D, Altman D, *et al*. A protocol of a cross-sectional study evaluating an online tool for early career peer reviewers assessing reports of randomised controlled trials. *BMJ Open* 2017;**7**:e017462. doi:10.1136/bmjopen-2017-017462

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-017462).

Received 2 May 2017
Revised 20 July 2017
Accepted 28 July 2017



For numbered affiliations see end of article.

**Correspondence to**
Dr Anthony Chauvin;
anthony.chauvin@aphp.fr

## ABSTRACT

**Introduction** Systematic reviews evaluating the impact of interventions to improve the quality of peer review for biomedical publications highlighted that interventions were limited and have little impact. This study aims to compare the accuracy of early career peer reviewers who use an innovative online tool to the usual peer reviewer process in evaluating the completeness of reporting and switched primary outcomes in completed reports.

**Methods and analysis** This is a cross-sectional study of individual two-arm parallel-group randomised controlled trials (RCTs) published in the BioMed Central series medical journals, *BMJ*, *BMJ Open* and *Annals of Emergency Medicine* and indexed with the publication type 'Randomised Controlled Trial'. First, we will develop an online tool and training module based (a) on the Consolidated Standards of Reporting Trials (CONSORT) 2010 checklist and the Explanation and Elaboration document that would be dedicated to junior peer reviewers for assessing the completeness of reporting of key items and (b) the Centre for Evidence-Based Medicine Outcome Monitoring Project process used to identify switched outcomes in completed reports of the primary results of RCTs when initially submitted. Then, we will compare the performance of early career peer reviewers who use the online tool to the usual peer review process in identifying inadequate reporting and switched outcomes in completed reports of RCTs at initial journal submission. The primary outcome will be the mean number of items accurately classified per manuscript. The secondary outcomes will be the mean number of items accurately classified per manuscript for the CONSORT items and the sensitivity, specificity and likelihood ratio to detect the item as adequately reported and to identify a switch in outcomes. We aim to include 120 RCTs and 120 early career peer reviewers.

**Ethics and dissemination** The research protocol was approved by the ethics committee of the INSERM Institutional Review Board (21 January 2016). The study is based on voluntary participation and informed written consent.

**Trial registration number** NCT03119376.

## Strengths and limitations of this study

- ► To the best of our knowledge, this will be the first study that will assess the efficacy of using an innovative online tool to help early career peer reviewers to assess the quality of reporting of randomised controlled trials.
- ► A randomised selection of trials from a large panel of biomedical journals.
- ► The absence of background for choosing the a priori effect size.

## INTRODUCTION

The peer review process is a cornerstone of biomedical research publication.[1] The editorial peer review process is described as journal editors relying on the views of independent experts in making decisions on, for example, the publication of submitted manuscripts or presentation of reports at meetings.[2] The peer review system is considered the best method for helping scientific editors decide on the acceptability of a manuscript for publication.[3] Nevertheless, the effectiveness of the system has been questioned. In 2010, a report commissioned by the UK House of Commons showed that the cost to the UK Higher Education Institutions in terms of staff time was £110 to £165 million per year for peer review and up to £30 million per year for the work done by editors and editorial boards.[4] Worldwide, peer review is estimated to cost £1.9 billion annually and accounts for about one-quarter of the overall costs of scholarly publishing and distribution.[5] The human resources were estimated to be about 15 million hours by 2010.[6]

Furthermore, the peer review system may fail to identify important flaws in the quality of submitted manuscripts and published articles.[7 8] A recent study by Hopewell *et al*[9] showed that peer reviewers failed to detect important deficiencies in the reporting of methods and results of randomised trials.

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



There is also considerable indirect evidence of the failure of peer review. Systematic reviews of clinical questions highlight a high prevalence of key information not being provided in peer-reviewed articles. Moreover, there is evidence to suggest that peer reviewers are not able to detect fraud,[10] mistakes[11 12] or spin.[13 14] Systematic reviews evaluating the impact of interventions to improve the quality of peer review for biomedical publications highlighted few randomised controlled trials (RCTs) and that interventions such as training have little or no impact.[15 16]

How the peer review process is currently organised can explain the difficulties encountered. A number of tasks are expected from peer reviewers when assessing reports of an RCT.[17] They have to assess the completeness of reporting, whether outcomes were switched from that reported in the trial registry, the choice of the comparator and intervention, the relevance of outcomes, the risk of bias, the external validity, the presence of spin, the originality, importance and relevance of the RCT to the journal's readers, etc. It may be unrealistic to expect a peer reviewer to complete all of these tasks.

Furthermore, the assessment of the completeness of reporting should be a prerequisite in the peer review process, because lack of transparency is a barrier to the assessment of important aspects of a trial such as the risk of bias[18] and reproducibility.[19] Another essential task of peer review is the identification of switched outcomes (ie, a discrepancy between the prespecified outcomes reported in the trial registry or protocol and outcomes reported in the manuscript). To avoid reporting bias(es),[20] all prespecified outcomes should be reported in the manuscript, and authors should report any changes to trial outcomes after the trial started, with reasons.[21] The Centre for Evidence-Based Medicine Outcome Monitoring Project (COMPare) assessed RCTs published in the five general medical journals with the highest impact factor. They compared the outcomes published to the outcomes recorded in trial registries or protocols. Among the 67 RCTs examined in July 2016 (www.COMPare-Trials.org), only nine had no discrepancies; 354 outcomes were not reported and 357 new outcomes were 'silently' added.[18]

Despite being essential, the assessment of the completeness of the reporting and the identification of switched outcomes are not appropriately performed.[9 22] These tasks do not require a high level of expertise and could be performed by researchers early on in their career. Early career peer reviewers could use a simple online tool based on the CONSORT 2010 checklist and Elaboration and Explanation publication[20] for reporting parallel-group RCTs developed according to the same principles used for the CobWeb tool.[23]

CobWeb is a writing aid tool based on CONSORT 2010. It was developed to provide guidance to authors when writing a manuscript of an RCT. The content of the tool was based on the CONSORT 2010 checklist and the Explanation and Elaboration document. For each domain, the tool comprised the corresponding CONSORT checklist item(s), bullet points with the key elements that need to be reported extracted from the Explanation and Elaboration document, as well as (an) example(s) of good reporting.

The peer reviewer tool will use the existing CobWeb elements reworded to be appropriate for peer reviewers and for the items not considered in CobWeb, we will extract the bullet points from the Explanation and Elaboration document of the CONSORT 2010 following the same process.[21]

The tool will feature bullet points eliciting the meaning of each checklist item. This tool will allow for standardising the completion of this task, be an opportunity to train early career peer reviewers to become more expert peer reviewers and be used to provide standardised and personalised feedback to authors based on the CONSORT items and bullet points; the authors would receive a request to report each item (with associated bullet points) that the early career peer reviewer identified as being inadequately reported.

The objectives of this project are as follows:

1. Develop an online tool and training module based (a) on the CONSORT 2010 checklist and the Explanation and Elaboration document modified for early career peer reviewers (ie, early stage researchers: masters students, PhD students, residents involved in clinical research during their study and clinicians involved in clinical research who never peer reviewed a manuscript) for assessing the completeness of reporting of 10 key items and (b) the COMPare process used to identify switched in primary outcomes in completed reports of RCTs.
2. Compare the performance of early career peer reviewers who use the peer reviewer tool with usual peer reviewers in identifying inadequate reporting and switched outcomes in completed reports of RCTs.

The planning, conduct, analysis and reporting of the study were discussed by the study's scientific committee.

## METHODS AND ANALYSIS

### Study design

We approached the issue of peer reviewing as if it were a diagnostic test aimed at detecting incomplete reporting. Before evaluating this 'diagnostic test' in an RCT, we aimed to evaluate its accuracy in a cross-sectional study. For this purpose, we defined a gold standard (ie, assessment by systematic reviewers) and two diagnostic tests to be compared: usual peer reviewers and early career peer reviewers trained and using the online tool.

This is a cross-sectional diagnostic study comparing the accuracy of early career peer reviewers who use the tool to that of usual peer reviewers when evaluating the completeness of reporting and a switched in primary outcome(s) in completed reports of RCTs at the first submission (figure 1). All procedures in the study will be consistent with ethical practice. This study received ethical approval from the Institutional Review Board of the INSERM ethics committee (IRB 17–335). The protocol was written

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



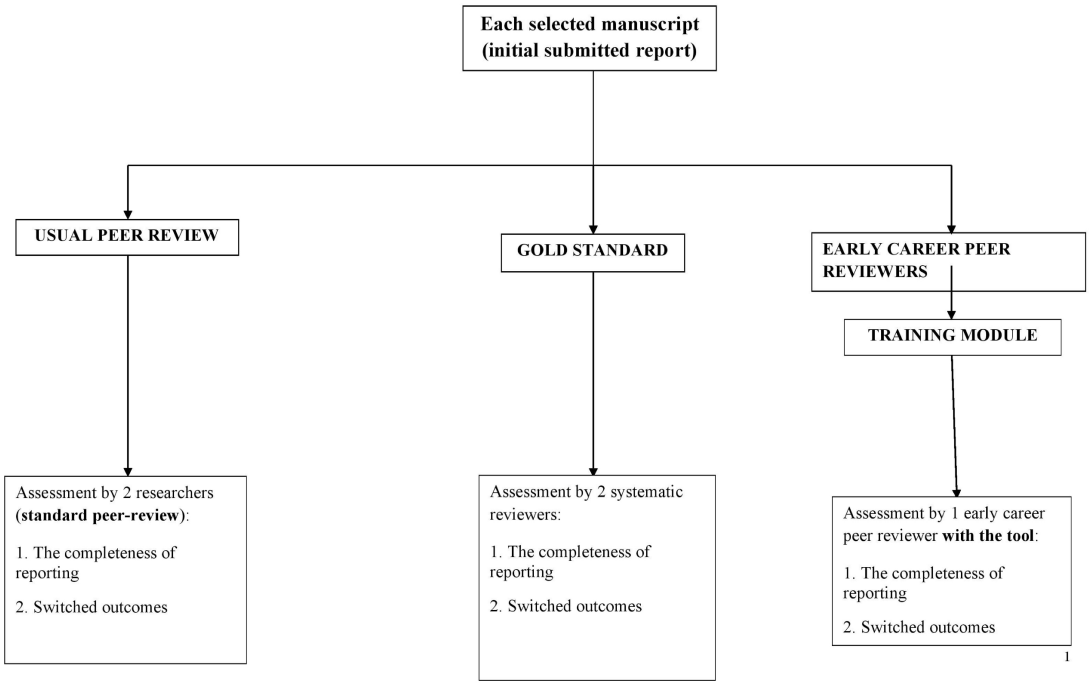



**Figure 1** The study design.

following the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidelines.

## The tool: COBPeer

### Principles of the tool

The online peer review tool provides guidance to early career peer reviewers when they are assessing the completeness of reporting of RCTs based on the relevant CONSORT 2010 checklist items, switched in primary outcomes from trial registry records based on the process used in COMPare and the associated CONSORT 2010 Elaboration and Explanation document.

Before using the tool, the early career peer reviewer will undergo an online training exercise consisting of evaluating extracts of manuscripts for each CONSORT item and corresponding trial registration records with specific feedback according to their answers.

The content of the tool and training will be tested and validated by the scientific committee of the study.

### Completeness of the reporting of RCTs section

To assess the completeness of reporting of RCTs, the tool will be developed according to the same principles used for developing the CobWeb tool.[23]

We will focus on the CONSORT items that are least often reported (ie, reported in less than 50% of the reports assessed as shown in the study published by Hopewell *et al*[9]) and are necessary to conduct a meta-analysis (ie, items of the risk of bias tool, outcomes data, safety data).

The selected items are:

- five items in the methods section (outcomes (6a), randomisation/sequence generation (8a), allocation concealment mechanism (9), blinding (11 a, 11b));
- four items in the results section (participant flow (13a, 13b), outcomes and estimation (17a), harms (19));
- one item (23) dedicated to trial registration (reporting registration number)

We decided to focus on these items because they are essential for systematic reviewers when evaluating the risk of bias and recording the outcome data.

Of note, we relied on the CONSORT extension for harm[24] to develop the bullet point for item 19 and on the SPIRIT guidelines explanatory document[25] to develop the bullet point for item 6a.

An item will be considered adequately reported if all bullet points related to this item are indicated as being reported. For example, item 11 'Blinding' will be considered adequately reported if the authors reported all the information below:

- Who (ie, participants, healthcare providers, data collectors, outcome adjudicators, and data analysts) was blinded to treatment assignments.
- How the blinding was performed (eg, used of placebo, intervention by physician unaware of the study).
- The similarities of the characteristics of the interventions (eg, appearance, taste, smell, method of administration).

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



The content of the tool is available in table 1.

### Switched outcomes section

The switched outcomes section will be based on the process used in COMPare that was simplified. The early career peer reviewer will have to:

1. find the trial's registry entry, by using http://google.com/;
2. copy the link to the online registry;
3. copy the date of the first registration;
4. copy the date of the beginning of the study;
5. copy the prespecified original primary outcome(s) from the registry;
6. compare the primary outcome(s) reported in the manuscript to the primary outcome(s) reported in the registry and determine whether there was a switch in outcome (ie, primary outcome added, deleted or changed);
7. indicate if authors justified the switched outcome(s) in the manuscript.

Overall the RCT manuscript will be classified as

- no switch in outcome;
- presence of switch in outcomes (at least one outcome was switched);
- unable to assess.

Of note, the early career peer reviewer will not search for the protocol or assess the protocol as was done in COMPare because of time constraints.

### Online training module

The online training module will start with a short reminder about the importance of the completeness of reporting and the assessment of switched outcomes from the trial register. For each item, the tool will provide bullet points of the item with explanations. Then, early career reviewers will assess passages from RCTs that are adequately or inadequately reported by using the tool. For each item assessed, participants will receive personalised feedback with a detailed explanation of the reported and missing bullet points.

## Participants

### Inclusion criteria

Early career peer reviewers are defined as early stage researchers: master students, PhD students, residents involved in clinical research during their study and clinicians who have never reviewed a manuscript.

### Exclusion criteria

Participants will be excluded if they have already peer reviewed one RCT.

### Setting and recruitment

We will use a large strategy of participants' recruitment. Participants will be identified from:

- editors of biomedical journals. We will contact specifically contact editors of journals interested in research in peer review field (eg, *BMC Medicine*, *Trials*, *BMC Peer Review and Research Integrity*);
- learnt societies (eg, European Society of Emergency Medicine);
- network of international students (eg, Students 4 Best Evidence);
- University Paris Descartes and Paris Diderot students.

We will send a standardised email to their contacts for promoting our study. The email will invite people to participate in an online training course on peer review. They will be informed of the inclusion criteria and about the different steps of the study (to complete an online training module and then peer review one manuscript randomly selected from our sample). If they agree to take part, they will then log onto the system.

### Blinding

Evaluators (systematic reviewers) will be blinded to the peer reviewers' assessment, early career peer reviewers' assessments and the content of the tool. They will have access to only one manuscript to assess. They will be instructed to base their assessment on only the content of the manuscript.

The evaluation of peer reviewers was blinded as performed before the study.

Early career peer reviewers will be blinded to the aim of the study, systematic reviewers' assessments and peer reviewers' comments. Systematic reviewers' assessments will not be accessible. Peer reviewers' comments in theory are accessible for articles published in BioMed Central series medical journals, *BMJ* and *BMJ Open*; however, the journal name, authors' names and study title will be masked to early career peer reviewers.

### Manuscript selection

We searched the US National Library of Medicine's PubMed database to identify all primary reports of two-arm parallel-group RCTs with an individual randomisation unit, whatever the treatment evaluated, that were published in the BioMed Central series medical journals, *BMJ, BMJ Open* and *Annals of Emergency Medicine* and indexed with the publication type 'Randomised Controlled Trial'. We included journals from BioMed Central series medical journals publishing at least 5 RCT reports per year between 1 January 2015 and 13 December 2016. We retrieved 1600 citations. One researcher screened all titles and abstracts and included all primary reports of RCTs assessing interventions in human participants who were randomly allocated to intervention groups. We excluded cluster RCTs, cross-over trials, equivalence and non-inferiority trials, feasibility studies, cost-effectiveness studies, phase I trials, study protocols, non-randomised controlled trials, secondary publications or analyses, pilot studies, systematic reviews, methodology studies and early phase studies. The screening process identified 222 eligible reports, among which we randomly selected 120 reports of RCTs. For each RCT, we retrieved the submitted manuscript and the first-round peer reviewers' comments (available online or requested from editors/publishers). Citations for 17 RCTs with documents not available were

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



**Table 1** The most important items of the Consolidated Standards of Reporting Trials (CONSORT) 2010 statement and bullet points that early career peer reviewers will assess

| CONSORT Item | CONSORT item | |
|---|---|---|
| Outcomes | Item 6a. Completely defined prespecified primary outcome measures, including how and when they were assessed i | Yes / No |
| | Was the primary outcome(s) clearly identify (eg, the primary/main outcome was pain)?<br>If no go to next section<br>If yes answer the following questions,<br>Please check if the author clearly report for the primary outcome(s):<br>▲ The variable of interest (eg, pain, all-cause mortality)<br>▲ ▲ How the outcome was assessed (eg, VAS, Beck Depression Inventory score, pain scale)<br>▲ ▲ The analysis metric (eg, change from baseline, final value, time to event)<br>▲ ▲ The summary measure for each study group (eg, mean, proportion with score≥2)<br>▲ ▲ Time point of interest for analysis (eg, 3 months)*NA if survival analysis<br>▲ Who assessed the outcome (eg, the patient, doctor, nurse, caregiver, other) | |
| Randomisation<br>Sequence generation | Item 8a. Method used to generate the random allocation sequence | |
| | Did the author report:<br>▲ The method of sequence generation (eg, a random number table or computerised random number generator, or other) | Yes / No |
| Allocation concealment | Item 9. Mechanism used to implement the random allocation sequence (eg, sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned | |
| | Did the author report:<br>▲ How the care provider enrolling patients was blinded to the next assignment in the sequence. Possible methods can rely on<br>▲ For centralised or 'third-party' assignment (ie, use of a central telephone randomisation system, automated assignment system)<br>▲ Having a third party prepare the randomisation list and hide the allocation assignment in advance via numbered identical bottles or sequentially numbered, sealed, opaque envelopes<br>▲ If the mechanism of the random allocation sequence is completely described but the sequence is not adequately concealed, please tick yes | Yes / No |
| Blinding | Was the study blinded yes/ no<br>▲ If yes go to 11a<br>▲ If no go to 13a<br>Item 11a. If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how<br>Item 11b. If relevant, description of the similarity of interventions | |
| | Did the author report:<br>▲ Who (ie, participants, healthcare providers, data collectors, outcome adjudicators, and data analysts) was blinded to treatment assignments?<br>▲ How was performed the blinding? (eg, used of placebo, intervention by physician unaware of the study)<br>▲ The similarities of the characteristics of the interventions (eg, appearance, taste, smell, method of administration)*NA | Yes / No |
| Participant flow | Item 13a. For each group, the numbers of participants who were randomly assigned, received intended treatment and were analysed for the primary outcome<br>Item 13b. For each group, losses and exclusions after randomisation, together with reasons | |

Continued

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



**Table 1** Continued

| CONSORT Item | CONSORT item | | | |
|---|---|---|---|---|
| | Did the authors report a flow chart | Yes | No | |
| | Did the author report in the flow chart or in the text:<br>▲ Number of participants randomised in each group<br>▲ Number of participants who received intended treatment in each group<br>▲ Number of participants did not receive the allocated treatment with reasons in each group<br>▲ Number of participants lost to follow-up with reasons in each group<br>▲ Number of participants who discontinued intervention with reasons in each group<br>▲ Number of participants analysed for the primary outcome in each group<br>▲ Number excluded from analysis with reasons in each group | Yes | No | |
| | Outcomes and estimation | Item 17a. For each primary outcome, results for each group, and the estimated effect size and its precision (such as 95% CI) | Did the author report for primary outcome: (answer yes if it is true for all primary outcomes)<br>▲ Result in each group (mean (SD) or nb of event/n)<br>▲ Difference in estimated effect between groups (eg, OR, risk ratio (RR), risk difference (RD), HR, difference in median survival time, mean difference (MD))<br>▲ Precision for difference between groups (eg, 95% CI) | |

Continued

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



**Table 1** Continued

| CONSORT Item | CONSORT item | | | |
|---|---|---|---|---|
| Yes | No | Harms | | Item 19. All important harms or unintended effects in each group |
| | Did the author report:<br>▲ How harms-related information was collected (eg, mode of data collection, timing, attribution methods) | Yes | No | |
| | ▲ ▲ For each group, participant withdrawals due to harm | | | |
| | ▲ ▲ Results in each group for each harms type with denominator (mean (SD) or nb of event/n) | | | |
| Registration | Item 23. Registration number and name of registry | | | |
| | Did the author report:<br>▲ The registration number | Yes | No | |
| Consistency between data registered and reported | | | | |
| Did authors report the same primary outcome in the register and manuscript (same variable, same metric, same time point) or was the primary outcome added, deleted, changed | | | | |
| If the study was not registered, please tick the box □ | | | | |
| If the study was registered, report<br>□ The link to the online registry:<br>□ Date of the registration:<br>□ Date of the start of the study: | | | | |
| Was the primary outcome(s) reported in the register or manuscript not sufficiently described to identify switch in outcomes? | | Yes | No | |
| Did you identified any outcome(s) reported by the authors as primary outcome(s) while not registered as such? | | Yes | No | |
| Did you identified any outcome(s) registered as primary outcome but not reported as such in the manuscript? | | Yes | Noi | |
| Did you identified any change in terms of time frame, metric or other information between the primary outcome(s) registered and reported in the manuscript? | | Yes | No | |
| If yes, please list the discrepancies: | | | | |
| – | | | | |
| – | | | | |
| Did the authors justify the switched outcome(s) in the manuscript? | | Yes | No | NA |
| | | | | |

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



replaced by 17 other randomly selected RCTs. Overall, we identified 120 manuscripts from 24 different journals.

## INTERVENTIONS

### Gold standard

The gold standard will be the assessment of the completeness of reporting by systematic reviewers of the 10 CONSORT items for the first manuscript submitted, (ie, before any changes were made as part of the peer review process). This assessment corresponds to the assessment of the risk of bias (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, selective reporting of outcomes) and the extraction of efficacy and harm outcome data. Pairs of systematic reviewers will independently extract data from eligible reports; any differences between reviewers will be resolved by discussion, with the involvement of an arbitrator if necessary. Reviewers involved in the data extraction will have expertise in the conduct of systematic reviews and will assess the completeness of reporting from the systematic-reviewer perspective. They will not have access to the tool to avoid being influenced by the tool. Consequently, the gold standard will rate items as inadequately reported only if the reporting is a real barrier to the conduct of a systematic review. This situation will ensure that the assessment of the early career peer reviewer who use the tool is identifying inadequate reporting that is not relevant.

The systematic reviewers will also systematically compare the outcomes reported in the manuscript and the outcomes reported in the registry and will document any discrepancies.

Overall, for the first manuscript, before any changes were made as part of the peer review process, each item will be classified as adequately reported (yes/no) for the gold standard.

### Online tool used by early career peer reviewers

If early career peer reviewers agree to participate, they will start the training module. If they successfully complete the training module, they will then download one randomly selected manuscript and assess at the first version submitted at the first round by using the online tool. They will be asked to perform this assessment in 30 min maximum.

There will be one early career peer reviewer per manuscript.

### Usual peer reviewer

The usual peer review process assessment of the completeness of reporting and switched outcomes will be assessed from the peer reviewers' reports. Two researchers will read all the peer review reports for the first round. The researchers will determine whether the peer reviewers raised some concerns about the completeness of reporting of the 10 CONSORT items considered and identified a switch in outcomes between the manuscript and the register. The assessment of all peer review reports for each manuscript will be combined (ie, the item will be rated as incompletely reported if at least one peer reviewer rated it as such). The two researchers will be blinded to the gold standard assessment.

## OUTCOMES

### Main outcome measure

The primary outcome will be the mean number of items accurately classified per manuscript initially submitted to the journal.

For the completeness of the reporting of RCTs section; 10 CONSORT items will be considered: Outcomes (6a), Randomisation/sequence generation (8a), Allocation concealment mechanism (9), Blinding (11 a, 11b), Participant flow (13a, 13b), Outcomes and estimation (17a), Harms (19) and one item (23) dedicated to trial registration.

Each item will be classified as 'adequately reported' (yes/no).

For the switched outcomes section, the manuscript will be classified as:

- no switch in primary outcome;
- presence of switch in primary outcomes (at least one outcome was switched);
- unable to assess.

### Secondary outcome measure

The secondary outcomes will be the mean number of items accurately classified per manuscript for the 10 CONSORT items and the sensitivity, specificity and likelihood ratio to detect the item as adequately reported and to identify switch in outcomes.

Because the gold standard may be less accurate than the tool used by early career peer reviewers in detecting inadequate reporting, we will systematically submit apparent false-positive results to experts identified by the scientific committee, who will indicate whether they believe the item is adequately reported (table 2).

## STATISTICS

### Sample size calculation

We will allow for detecting an effect size of 0.3 for the mean number of items accurately classified per manuscript with a power of 90% and a two-sided alpha-level of 5%.

Each manuscript will be assessed by one early career peer reviewer; each early career peer reviewer will assess one manuscript. So, we will randomly select 120 reports of RCTs to be assessed in the study. We do not have any data on the mean (SD) number of items accurately classified by usual peer reviewers. For each manuscript, we will have a single assessment related to the usual peer review process that will combine the first assessment of all peer reviewers for that manuscript.

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



Table 2 Planned tabulations

| | | Gold standard | |
|---|---|---|---|
| | | Item completely reported | Item incompletely reported |
| Early career peer reviewers | Item completely reported | True positive | False positive |
| | Item incompletely reported | False negative | True negative |
| Usual peer reviewers | Item completely reported | True positive | False positive |
| | Item incompletely reported | False negative | True negative |

Item adequately classified=True positive+true negative.
The primary outcome will be the mean number of items accurately classified per manuscript across all the trial reports that are assessed.

### Statistical analysis

The primary outcome will be the mean number of items accurately classified per manuscript by usual peer reviewers and by early career peer reviewers, the gold standard being the systematic reviewers' assessment. We will compare the mean number of items accurately classified per manuscript by early career peer reviewers versus usual peer reviewers. This will be performed using a paired t-test. The sensitivity and specificity comparing the early career peer reviewers and the usual peer-reviewers' assessment will be determined by a test for pair proportion.

### Data management and monitoring

Data will be entered on at a secure, password-protected, website. As part of the data protection, confidentiality and anonymity requirements, all participants will be assigned a unique individual identifier. The database will be managed by the main investigators (CA and BI). Access will be limited to individuals deemed appropriate. No interim analyses will be performed.

## ETHICS AND DISSEMINATION

Authorisation by the Commission Nationale de l'Informatique et des Libertés (file number 2021376) whose purpose is to protect participants' personal data and the institutional review board of the INSERM ethics committee (IRB 17–335) (see Appendix 1 in the online supplementary file 1) was obtained and the study protocol was registered at ClinicalTrials.gov (http://clinicaltrials.gov, NCT03119376). Any changes or additions to the protocol will be registered on ClinicalTrials.gov. Participation will be voluntary. All participants will provide informed consent and may withdraw at any time (see Appendix 2 in the online supplementary file 1). Following their participation, they will receive one Continued Medical Education credit.

The results will be presented at international congresses, published in a peer-reviewed journal and summarised on ClinicalTrials.gov. In addition, the findings will be disseminated via an open access website so that they can be used by the wider community. Study results will also be disseminated directly to participants.

Any changes or additions to the protocol will be registered on Clinicaltrials.gov. Authorship of papers arising from this study will be based on contributions to the study including intellectual content as recommended by the International Committee of Medical Journal Editors.

The data set will be the property of the scientific committee. The results will be presented at international congresses, published in a peer-reviewed journal and summarised on ClinicalTrials.gov

## DISCUSSION

Several researchers have highlighted that the quality of reporting of peer-reviewed and editor-approved clinical research is poor.[26–28] However, because biomedical journals remains the best method for disseminating biomedical knowledge, it is fundamental to assess the implementation of novel interventions and must be assessed in different actors of the publication system to improve its efficiency and the quality of reporting. In this study, we limited our research to interventions to improve quality of reporting through peer review. To the best of our knowledge, this will be the first study of an innovative online tool to help early career peer reviewers to assess the quality of reporting of RCTs.

**Author affiliations**
[1]METHODS team, Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité (CRESS), Paris, France
[2]Paris Descartes University, Paris, France
[3]Department of Emergency, Hôpital Lariboisière, Assistance Publique des Hôpitaux de Paris, University Diderot, Paris, France
[4]Centre for Journalology, Ottawa Hospital Research Institute, Ottawa, Canada
[5]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada
[6]Centre for Statistics in Medicine, University of Oxford, Oxford, UK
[7]Department of Emergency Medicine, School of Medicine, University of California, Los Angeles, California, USA
[8]Faculty of 1000, London, UK
[9]Oxford Clinical Trials Research Unit, University of Oxford, Oxford, UK
[10]BioMed Central Ltd, London, UK
[11]METHODS team, Centre d'Épidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique des Hôpitaux de Paris, Paris, France

**Contributors** AC and IB were responsible for the concept of the study and conceived and designed the study. DM, DGA, DS, SA, SH, DRS and PR helped to write the draft protocol and made a critical contribution to the content. All authors read and approved the final protocol.

**Competing interests** None declared.

**Patient consent** Obtained.

**Ethics approval** The Institutional Review Board of INSERM ethics committee (IRB 17-335).

Downloaded from http://bmjopen.bmj.com/ on September 27, 2017 - Published by group.bmj.com



**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** All data will be published.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

1. Kassirer JP, Campion EW. Peer review. Crude and understudied, but indispensable. *JAMA* 1994;272:96–7.
2. Jefferson T, Rudin M, Brodney Folse S, *et al*. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Syst Rev* 2007:MR000016.
3. Rennie R. Editorial peer review: its development and rationale. In: Godlee F, Jefferson T, eds. *Peer review in health sciences*. Second edition. London: BMJ Books, 2003:1––13.
4. Public Library of Science. *"Peer review—optimizing practices for online scholarly communication," in Peer Review in Scientific Publications, Eighth Report of Session 2010–2012, Vol. I: Report, Together with Formal, Minutes, Oral and Written Evidence, eds House of Commons Science and Technology Committee, editor*. London: The Stationery Office Limited, 2011:p21–p22.
5. Public Library of Science. *"Peer review—optimizing practices for online scholarly communication," in Peer Review in Scientific Publications, Eighth Report of Session 2010–2012, Vol. I: Report, Together with Formal, Minutes, Oral and Written Evidence, eds House of Commons Science and Technology Committee, editor*. London: The Stationery Office Limited, 2011:p174–p178.
6. Association of American Publishers. Digital licenses replace print prices as accurate reflection of real journal costs, 2012. http://publishers.org/sites/default/files/uploads/PSP/summer-fall_2012.pdf. (Accessed 08 june 2016).
7. Ghimire S, Kyung E, Kang W, *et al*. Assessment of adherence to the CONSORT statement for quality of reports on randomized controlled trial abstracts from four high-impact general medical journals. *Trials* 2012;13:77.
8. Boutron I, Dutton S, Ravaud P, *et al*. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.
9. Hopewell S, Collins GS, Boutron I, *et al*. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ* 2014;349:g4145.
10. Bohannon J. Who's afraid of peer review? *Science* 2013;342:60–5.
11. Schroter S, Black N, Evans S, *et al*. What errors do peer reviewers detect and does training improve their ability to detect them? . *J R Soc Med* 2008;101:507–14.
12. Walbot V. Are we training pit bulls to review our manuscripts? *J Biol* 2009;8:24.
13. Lazarus C, Haneef R, Ravaud P, *et al*. Peer reviewers identified spin in manuscripts of nonrandomized studies assessing therapeutic interventions, but their impact on spin in abstract conclusions was limited. *J Clin Epidemiol* 2016;77:44–51.
14. Stahel PF, Moore EE. Peer review for biomedical publications: we can improve the system. *BMC Med* 2014;12:179.
15. Bruce R, Chauvin A, Trinquart L, *et al*. Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. *BMC Med* 2016;14:85.
16. Galipeau J, Moher D, Campbell C, *et al*. A systematic review highlights a knowledge gap regarding the effectiveness of health-related training programs in journalology. *J Clin Epidemiol* 2015;68:257–65.
17. Chauvin A, Ravaud P, Baron G, *et al*. 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 Med* 2015;13:158.
18. COMPare-trials.org. Tracking switched outcomes in clinical trials. http://compare-trials.org/. (accessed 2 May 2016).
19. Munafò M. Metascience: reproducibility blues. *Nature* 2017;543:619–20.
20. Chan AW, Hróbjartsson A, Haahr MT, *et al*. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457–65.
21. Moher D, Hopewell S, Schulz KF, *et al*. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
22. Mathieu S, Chan AW, Ravaud P. Use of trial register information during the peer review process. *PLoS One* 2013;8:e59910.
23. Barnes C, Boutron I, Giraudeau B, *et al*. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. *BMC Med* 2015;13:221.
24. Ioannidis JP, Evans SJ, Gøtzsche PC, *et al*. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781–8.
25. Chan AW, Tetzlaff JM, Altman DG, *et al*. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:200–7.
26. Yordanov Y, Dechartres A, Porcher R, *et al*. Avoidable waste of research related to inadequate methods in clinical trials. *BMJ* 2015;350:h809.
27. Dwan K, Altman DG, Clarke M, *et al*. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* 2014;11:e1001666.
28. Page MJ, Shamseer L, Altman DG, *et al*. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13:e1002028.

## 5.10 Appendices de l'article

Appendix 1. The institutional review board authorization

**CEEI - IRB**

**Comité d'Evaluation Ethique de l'Inserm**

**IRB00003888**

Instituts thématiques — Inserm — institut national de la santé et de la recherche médicale

**Nos réf: CD / VB 17-008**
Dossier suivi par :
Christine DOSQUET – CEEI
@ : ceei@inserm.fr

**Monsieur Anthony CHAUVIN**
**Centre d'Epidemiologie Clinique**
**Hôpital Hôtel Dieu**
**1, place du Parvis Notre-Dame**
**75181 PARIS Cedex 4**

Paris, January 21st 2017

**To whom it may concern**
Opinion number 17-355

**Sir,**

The Ethics Evaluation Committee of Inserm, the Institutional Review Board (IRB00003888, IORG0003254, FWA00005831) of the French Institute of Health and Medical Research, has reviewed and approved the research project entitled:

**« Development and evaluation of an online tool dedicated to junior peer reviewers for assessing the reporting of randomized controlled trial reports: a protocol ».**

The investigator undertakes to respect the protocol and to follow the recommendations proposed by the ethics evaluation committee.

Yours sincerely,

Christine DOSQUET
IRB President

Appendix 2. Information for participants

**INFORMATION FOR PARTICIPANTS**

**Title: Online Training Moduleto Peer Review.**

Dear colleague /XX/student,

We would like to invite you to participate in an academic study of an online training module and a specific tool dedicated to early career peer reviewers when assessing a randomized controlled trial.

This study is performed in collaboration with medical journal editors, the CONSORT group and the EQUATOR network.

When the study is completed, you will receive 1 credit of Continued Medical Education (CME) for your participation in this training programme. Further, we will send you the study results and acknowledge your participation, if you agree to have your name cited in the manuscript acknowledgments section. Finally, this programme is a unique opportunity for you to gain expertise in peer-reviewing and become a peer reviewer for journal editors.

This training programme is focused on the assessment of the completeness of reporting in manuscripts of randomized controlled trials according to the CONSORT statements, and the switch in primary outcomes, which are essential elements in the peer review process. During this programme, you will 1) assess the completeness of reporting of manuscripts extracts using the tool (with immediate feedback and explanation of incorrect responses), and 2) assess a full text article using the tool.

If you are successful (i.e. >80% appropriate answers), you will have to assess a final manuscript.

We estimate that this process with take about one hour.

All information collected during this study will be treated confidentially.

Thank you very much in advance for your help with this project

Sincerely,

Pr Isabelle Boutron (Paris Descartes University, INSERM, France),
Dr Anthony Chauvin (Paris Descartes University, INSERM, France)
Dr David Moher (Ottawa Hospital Research Institute, Canada)
Pr Philippe Ravaud (Paris Descartes University, INSERM, France),

*Authorization by CNIL ("Commission National de l'Information et des libertés") whose remit is to protect participants' personal data and the institutional review board of INSERM ethics committee (IRB xxxxxx) was obtained. In accordance with the law "Informatique et Libertés" of 6 January 1978 amended in 2004, you have the right to access and rectify any information concerning you, which can be exercised by contacting Professor Isabelle Boutron, Research Center Epidemiology and statistics Sorbonne Paris Cité (isabelle.boutron@htd.aphp.fr).*

*If you have any questions during your participation in this study, you can contact the person in charge of the study, Professor Isabelle Boutron, tel: +33 (0)1 42 34 78 33, mail: isabelle.boutron@htd.aphp.fr*

# 6. Discussion et perspectives

## 6.1 Principaux résultats

Le processus de peer-review est considéré comme la pierre angulaire de la publication scientifique. Or, des travaux ont montré que les évaluations des peer-reviewers sont souvent discordantes, qu'ils ne parviennent pas à détecter la fraude, les erreurs et que leur évaluation est influencée par les résultats. Face au coût humain et financier de ce processus, il est primordial de l'étudier et de l'implémenter pour améliorer son efficience.

Ce travail de thèse a permis : 1) d'identifier les tâches attendues des peer-reviewers lors de l'évaluation d'un manuscrit rapportant un essai contrôlé randomisé, et de comparer l'importance de ces tâches selon le point de vue des peer-reviewers et des éditeurs, 2) d'identifier les interventions visant à améliorer la peer-review et ayant démontré leur efficacité, 3) de développer et évaluer l'impact d'un outil informatisé permettant de standardiser les critiques des reviewers sur la qualité de la présentation des articles dédié aux peer-reviewers juniors.

Le premier travail a permis de montrer qu'un grand nombre de tâches étaient demandées aux peer-reviewers lors de l'évaluation d'un manuscrit. De plus, les tâches les plus importantes pour les peer-reviewers n'étaient pas conformes aux tâches les plus souvent demandées par les éditeurs. Ce premier travail a permis de mettre en lumière un des motifs de la faible qualité de la peer-review. Ce travail a été cité dans plusieurs éditoriaux (143) et a été cité 11 fois depuis sa publication en 2015.

Dans le deuxième travail, nous avons identifié 22 essais contrôlés randomisés, rapportant 25 comparaisons évaluant 5 interventions : l'entrainement/la formation, l'ajout d'une peer-review spécifique, l'utilisation de checklist, la peer-review en ouvert et la peer-review en aveugle. Seules deux interventions ont montré un impact positif : l'ajout d'une évaluation par un statisticien permettant d'améliorer la qualité du manuscrit final et l'évaluation en ouvert permettant

d'améliorer la qualité du rapport des peer-reviewers. Ce deuxième travail a mis en évidence le faible nombre d'essais contrôlés randomisés réalisés pour améliorer la qualité de la peer-review. Ce travail a permis de sensibiliser la communauté scientifique sur ce manque crucial de recherche dans le domaine de la peer-review (144) (145) (146). De plus, des auteurs ont proposé des interventions originales visant à automatiser le processus de peer-review (146) (147) (148).

Dans le dernier travail nous avons développé un outil en ligne (COBPeer) d'aide à l'évaluation de la qualité de la présentation des essais contrôlé randomisés. L'utilisation de celui-ci par des jeunes chercheurs semble améliorer l'efficience de la peer-review grâce à une meilleure détection des items mal rapportés par les auteurs (150).

## 6.2 Discussion

Dans nos travaux, nous nous sommes focalisés seulement sur les peer-reviewers. Or le processus de peer-review est un éco système complexe avec de nombreux acteurs : auteurs, éditeurs et peer-reviewers. Chaque acteur a un déterminisme propre, mais aussi des interactions influençant les décisions propres de chaque acteur. Ainsi, il est important de repenser la peer-review à deux niveaux : macroscopique évidemment, mais aussi microscopique.

Actuellement au niveau microscopique, c'est-à-dire au niveau du manuscrit soumis, celui-ci va être évalué par un nombre défini de peer-reviewers ayant chacun une expérience hétérogène. Or tous les peer-reviewers ont un même objectif : l'évaluation qualitative du manuscrit. Comme nous l'avons déjà évoqué, cet objectif ne signifie pas la même chose pour chaque peer-reviewer. Ainsi, repenser la peer-review signifie avant tout redéfinir clairement ses objectifs. La re-définition de ces objectifs pourrait dans un premier temps aider à repenser le processus de peer-review. A savoir, ce processus pourrait consister en utilisation d'un maximum de peer-reviewers avec une spécificité très pointue pour évaluer une petite partie de l'article (i.e., statisticiens,

méthodologistes, linguistes). Ou à l'inverse, augmenter drastiquement le nombre de peer-reviewers pour maximiser la probabilité de trouver des erreurs dans le manuscrit. Bien entendu, ces deux systèmes, que l'on pourrait qualifier de simplistes, se heurtent à l'écueil du nombre de peer-reviewers disponibles ou disposés à évaluer un manuscrit. Des auteurs ont mis en évidence que la recherche sur l'amélioration de l'efficience de la peer-review ne doit pas signifier une complexification du processus (151). Car plus le processus serait complexe et contraignant, moins il y aurait de peer-reviewers.

Kovanis et al. ont publié en 2017 (148,152) une étude dans laquelle ils ont considéré le processus de peer-review comme un ensemble macroscopique constitué de différentes identités : les éditeurs, auteurs et les peer-reviewers. Les auteurs ont modélisé 5 types de peer-review différentes : système à publication immédiate avec ou sans relecture en ligne (i.e., *crowdsourcing*), système avec un seul tour de peer-review autorisée (i.e., *re-review opt-out*), systèmes de partage de revues dans lesquels les manuscrits rejetés sont resoumis avec leurs révisions à d'autres journaux (i.e., *portable*) ou à des journaux du même éditeur mais de moindre impact (i.e., *cascade*). Les auteurs ont mesuré la performance de la peer-review, la durée du processus de peer-review et la dissémination des articles. Les deux systèmes avec publication immédiate avaient permis d'améliorer la qualité des publications, mais n'ont pas fait gagner de temps ni augmenté le nombre de manuscrit publiés par rapport au processus consensuel de peer-review. Le système avec un seul tour de peer-review diminuait le temps du processus mais échouait à identifier les articles à ne pas publier. Cela avait un impact direct sur la performance qualitative de la peer-review. Le modèle *cascade* semblait améliorer l'efficience de la peer-review tout en diminuant le temps du processus de peer-review et en favorisant la dissémination de la recherche.

## 6.3 Perspectives

A la suite de nos travaux, il semble pertinent de poursuivre la réflexion sur la peer-review.

### 6.3.1 Une peer-review en 2 étapes

Suite à notre dernier travail, nous souhaitons implémenter le processus de peer-review de journaux biomédicaux avec une peer-review en deux étapes.

Nos travaux ont montré le nombre important de tâches demandées aux peer-reviewers. Comme nous l'avons vu, ces tâches concernent aussi bien la forme que le fond. Ainsi, notre outil COBPeer pourrait potentiellement permettre d'améliorer l'efficience de la peer-review. En effet, rajouter une première étape avec des peer-reviewers juniors évaluant l'article soumis permettrait permettrait de créer un filtre « qualitatif » de manuscrits soumis et envoyés aux peer-reviewers experts. Le temps ainsi passé à la peer-review par les experts sera moindre car ceux-ci ne devront plus que s'intéresser qu'à la question de la recherche, ainsi que sur le fond et non sur la forme.

Pour évaluer cette proposition, il serait intéressant d'implémenter notre outil dans le processus de peer-review.

Ainsi, chaque essai contrôlé randomisé serait randomisé en deux bras : un bras avec une peer-review en 2 étapes et un bras contrôle avec une peer review classiques. Les critères de jugement principaux seraient le nombre moyen d'items du CONSORT correctement rapportés par article et la durée du processus de peer-review. Le critère de jugement secondaire serait le taux de rejet.

### 6.3.2 Autres formes de peer-review

Face à l'émergence de l'intelligence artificielle dans la recherche biomédicale, de nombreux auteurs cherchent à évaluer son apport dans le processus de peer-review. Le développement

d'outil informatique permettant la recherche automatique de données dans l'article pourrait permettre une automatisation de la peer-review.

Ce type de peer-review automatique est en cours d'évaluation au sein de certain journaux du groupe BioMed Central (153). Un outil informatique *StatReviewer* permet à la fois d'évaluer si les auteurs rapportent l'ensemble des items du CONSORT mais aussi d'évaluer la pertinence des analyses statistiques réalisées par les auteurs.

Des chercheurs ont quant à eux cherché à automatiser l'évaluation du risque de biais des essais randomisés (154). Pour ce faire, ils ont créé un système informatique du nom de RobotReviewer dont l'objectif était d'évaluer le risque de biais et de le justifier avec des extraits de l'article. Les auteurs ont fait évaluer à leur système 12 808 articles ayant étaient évalué au décours d'une revue systématique de la Cochrane. Les auteurs ont intégré dans leur analyse le fait que l'évaluation du risque de biais pour un article donné pouvait varier d'un évaluateur à l'autre (155).

Le critère de jugement était la concordance entre l'évaluation du risque de biais par des relecteurs de la Cochrane et celle de RobotReviewer ainsi que la pertinence de l'extrait sélectionné. RobotReviewer était moins précis dans le jugement du risque de biais mais la différence était inférieure à 10% (71% versus 78,3%, $p<0,01$). Par contre, la justification du texte était plus souvent considérée comme très pertinente dans une proportion équivalente à celle donnée par les peer-reviewers de la Cochrane (60,4% versus 56,5%). Les auteurs de RobotReviewer sont actuellement en train d'améliorer leur système (156).

Ce type d'évaluation pourrait permettre non seulement de gagner du temps mais aussi de standardiser l'évaluation des essais contrôlés randomisés et ainsi limiter l'hétérogénéité de la peer-review actuelle.

# 7. Conclusion

La peer-review est un système qui doit apprendre à se réinventer et se redéfinir. Des recherches supplémentaires sont nécessaires pour identifier les interventions qui pourraient améliorer le processus et évaluer ces interventions dans des essais de bonne qualité. L'apport des nouvelles technologies, telle que l'intelligence artificielle, pourrait être une réponse mais nécessite une évaluation précise.

## Bibliographie


1. Jefferson T, Rudin M, Brodney Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. Cochrane Database Syst Rev. 2007 Apr 18;(2):MR000016.
2. Kassirer JP, Campion EW. Peer review. Crude and understudied, but indispensable. JAMA. 1994 Jul 13;272(2):96–7.
3. Guthrie S, Ghiga I, Wooding S. What do we know about grant peer review in the health sciences? F1000Research. 2017;6:1335.
4. Kronick DA. Peer Review in 18th-Century Scientific Journalism. JAMA. 1990 Mar 9;263(10):1321–2.
5. Spier R. The history of the peer-review process. Trends Biotechnol. 2002 Aug 1;20(8):357–8.
6. Spier RE, Bird SJ. Scientific misconduct: ongoing developments. Sci Eng Ethics. 2000 Jan;6(1):3–4.
7. Doubeni CA. The Assessment of Impact, and Scientific Premise and Rigor of Applications by Grant Review Panels. Epidemiol Camb Mass. 2016 May;27(3):314–5.
8. Eyler JM. The changing assessments of John Snow's and William Farr's cholera studies. Soz Praventivmed. 2001;46(4):225–32.
9. Burnham JC. The Evolution of Editorial Peer Review. JAMA. 1990 Mar 9;263(10):1323–9.
10. Rennie D. Editorial Peer Review in Biomedical Publication: The First International Congress. JAMA. 1990 Mar 9;263(10):1317–1317.
11. Rennie D, Flanagin A. The Second International Congress on Peer Review in Biomedical Publication. JAMA. 1994 Jul 13;272(2):91–91.
12. Rennie D. Let's make peer review scientific. Nat News. 2016 Jul 7;535(7610):31.
13. Csiszar A. Peer review: Troubled from the start. Nat News. 2016 Apr 21;532(7599):306.
14. Giles J. Journals submit to scrutiny of their peer-review process. Nature. 2006 Jan 19;439(7074):252.
15. Jefferson T, Wager E, Davidoff F. Measuring the quality of editorial peer review. JAMA. 2002 Jun 5;287(21):2786–90.
16. Gibson TA. Post-publication review could aid skills and quality. Nature. 2007 Jul 26;448(7152):408.
17. F1000 – Faculty of 1000. Available from: https://f1000.com/ Accessed 15 Jun 2018
18. PubPeer - Search publications and join the conversation. [Internet]. [cited 2018 Mar 2]. Available from: https://pubpeer.com/
19. ScienceOpen. Available from: https://www.scienceopen.com/ Accessed 15 Jun 2018
20. Eyre-Walker A, Stoletzki N. The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. PLoS Biol. 2013 Oct;11(10):e1001675.
21. Hunter J. Post-publication peer review: opening up scientific conversation. Front Comput Neurosci. 2012;6:63.
22. Teixeira da Silva JA, Dobránszki J. Problems with traditional science publishing and finding a wider niche for post-publication peer review. Account Res. 2015;22(1):22–40.
23. Bastian H. A stronger post-publication culture is needed for better science. PLoS Med. 2014 Dec;11(12):e1001772.

24. Post-publication criticism is crucial, but should be constructive. Nat News. 2016 Dec 1;540(7631):7.
25. Knoepfler P. Reviewing post-publication peer review. Trends Genet TIG. 2015 May;31(5):221–3.
26. Public Library of Science. Peer review—optimizing practices for online scholarly communication. In: House of Commons Science and Technology Committee, editor. Peer Review in Scientific Publications, Eighth Report of Session 2010–2012, Vol. I: Report, Together with Formal, Minutes, Oral and Written Evidence. London: The Stationery Office Limited; 2011. p. 174–8.
27. Rubriq. How We Found 15 Million Hours of Lost Time. In Rubriq: independent peer review system. Available from: https://www.aje.com/en/arc/peer-review-process-15-million-hours-lost-time/ Accessed 28 Jun 2018
28. Rowe BH, Strome TL, Spooner C, Blitz S, Grafstein E, Worster A. Reviewer agreement trends from four years of electronic submissions of conference abstract. BMC Med Res Methodol. 2006 Mar 19;6:14.
29. Jackson JL, Srinivasan M, Rea J, Fletcher KE, Kravitz RL. The validity of peer review in a general medicine journal. PloS One. 2011;6(7):e22475.
30. Weiner BK, Weiner JP, Smith HE. Spine journals: is reviewer agreement on publication recommendations greater than would be expected by chance? Spine J Off J North Am Spine Soc. 2010 Mar;10(3):209–11.
31. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. J Clin Oncol Off J Am Soc Clin Oncol. 2014 Dec 20;32(36):4120–6.
32. Lazarus C, Haneef R, Ravaud P, Boutron I. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. BMC Med Res Methodol. 2015 Oct 13;15:85.
33. Amos AJ. A review of spin and bias use in the early intervention in psychosis literature. Prim Care Companion CNS Disord. 2014;16(1).
34. Ghimire S, Kyung E, Kang W, Kim E. Assessment of adherence to the CONSORT statement for quality of reports on randomized controlled trial abstracts from four high-impact general medical journals. Trials. 2012 Jun 7;13(1):77.
35. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes. JAMA. 2010 May 26;303(20):2058–64.
36. Hopewell S, Collins GS, Boutron I, Yu L-M, Cook J, Shanyinde M, et al. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. BMJ. 2014 Jul 1;349:g4145.
37. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. JAMA. 2006 Mar 8;295(10):1147–51.
38. Chen Y-P, Chen L, Li W-F, Lee AWM, Vermorken JB, Wee J, et al. Reporting Quality of Randomized, Controlled Trials Evaluating Combined Chemoradiotherapy in Nasopharyngeal Carcinoma. Int J Radiat Oncol Biol Phys. 2017 01;98(1):170–6.
39. Agha RA, Fowler AJ, Limb C, Whitehurst K, Coe R, Sagoo H, et al. Impact of the mandatory implementation of reporting guidelines on reporting quality in a surgical journal: A before and after study. Int J Surg Lond Engl. 2016 Jun;30:169–72.
40. Rutterford C, Taljaard M, Dixon S, Copas A, Eldridge S. Reporting and methodological quality of sample size calculations in cluster randomized trials could be

improved: a review. J Clin Epidemiol. 2015 Jun;68(6):716–23.
41. Walleser S, Hill SR, Bero LA. Characteristics and quality of reporting of cluster randomized trials in children: reporting needs improvement. J Clin Epidemiol. 2011 Dec;64(12):1331–40.
42. Schroter S, Glasziou P, Heneghan C. Quality of descriptions of treatments: a review of published randomised controlled trials. BMJ Open. 2012;2(6).
43. Ioannidis JP, Caplan AL, Dal-Ré R. Outcome reporting bias in clinical trials: why monitoring matters. BMJ. 2017 Feb 14;356:j408.
44. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA. 2009 Sep 2;302(9):977–84.
45. Tracking switched outcomes in clinical trials [Internet]. COMPare. [cited 2018 Mar 1]. Available from: http://compare-trials.org
46. Howard B, Scott JT, Blubaugh M, Roepke B, Scheckel C, Vassar M. Systematic review: Outcome reporting bias is a problem in high impact factor neurology journals. PloS One. 2017;12(7):e0180986.
47. Lancee M, Lemmens CMC, Kahn RS, Vinkers CH, Luykx JJ. Outcome reporting bias in randomized-controlled trials investigating antipsychotic drugs. Transl Psychiatry. 2017 Sep 12;7(9):e1232.
48. Wiebe J, Detten G, Scheckel C, Gearhart D, Wheeler D, Sanders D, et al. The heart of the matter: Outcome reporting bias and registration status in cardio-thoracic surgery. Int J Cardiol. 2017 Jan 15;227:299–304.
49. van Lent M, IntHout J, Out HJ. Differences between information in registries and articles did not influence publication acceptance. J Clin Epidemiol. 2015 Sep;68(9):1059–67.
50. Schroter S, Black N, Evans S, Godlee F, Osorio L, Smith R. What errors do peer reviewers detect, and does training improve their ability to detect them? J R Soc Med. 2008 Oct 1;101(10):507–14.
51. Gasparyan AY, Ayvazyan L, Akazhanov NA, Kitas GD. Conflicts of interest in biomedical publications: considerations for authors, peer reviewers, and editors. Croat Med J. 2013 Dec;54(6):600–8.
52. Resnik DB, Elmore SA. Conflict of Interest in Journal Peer Review. Toxicol Pathol. 2018 Feb 1;46(2):112–4.
53. Viswanathan M, Carey TS, Belinson SE, Berliner E, Chang S, Graham E, et al. Identifying and Managing Nonfinancial Conflicts of Interest for Systematic Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality. 2013 (AHRQ Methods for Effective Health Care). Available from: http://www.ncbi.nlm.nih.gov/books/NBK148586/
54. Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. BMJ. 2002 Aug 3;325(7358):249.
55. WAME. World association of medical editors. Available from: http://www.wame.org/ Accessed 29 Jun 2018
56. WANE. Conflict of Interest in Peer-Reviewed Medical Journals. Available from: www.wame.org/about/conflict-of-interest-in-peer-reviewed-medical Accessed 29 Jun 2018
57. Committee on publication ethics? Code of conduct and best practice guidelines for journal editors. Available from: http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf Accessed 29 Jun 2018
58. Anraku A, Jin Y-P, Trope GE, Buys YM. Survey of conflict-of-interest disclosure policies of ophthalmology journals. Ophthalmology. 2009 Jun;116(6):1093–6.

59. Simera I, Altman DG. Reporting medical research. Int J Clin Pract. 2013 Aug;67(8):710–6.
60. Roseman M, Milette K, Bero LA, Coyne JC, Lexchin J, Turner EH, et al. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. JAMA. 2011 Mar 9;305(10):1008–17.
61. Roseman M, Turner EH, Lexchin J, Coyne JC, Bero LA, Thombs BD. Reporting of conflicts of interest from drug trials in Cochrane reviews: cross sectional study. BMJ. 2012 Aug 16;345:e5155.
62. Hojat M, Gonnella JS, Caelleigh AS. Impartial judgment by the 'gatekeepers' of science: fallibility and accountability in the peer review process. Adv Health Sci Educ Theory Pract. 2003;8(1):75–96.
63. Kilwein JH. Biases in medical literature. J Clin Pharm Ther. 1999 Dec;24(6):393–6.
64. Grod ON, Budden AE, Tregenza T, Koricheva J, Leimu R, Aarssen LW, et al. Systematic variation in reviewer practice according to country and gender in the field of ecology and evolution. PloS One. 2008 Sep 12;3(9):e3202.
65. Lloyd ME. Gender factors in reviewer recommendations for manuscript publication. J Appl Behav Anal. 1990;23(4):539–43.
66. Gilbert JR, Williams ES, Lundberg GD. Is there gender bias in JAMA's peer review process? JAMA. 1994 Jul 13;272(2):139–42.
67. Hargens LL. Variation in journal peer review systems. Possible causes and consequences. JAMA. 1990 Mar 9;263(10):1348–52.
68. Tricco AC, Thomas SM, Antony J, Rios P, Robson R, Pattani R, et al. Strategies to Prevent or Reduce Gender Bias in Peer Review of Research Grants: A Rapid Scoping Review. PloS One. 2017;12(1):e0169718.
69. ICMJE. Responsibilities in the Submission and Peer-Review Process. Available from: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/responsibilities-in-the-submission-and-peer-peview-process.html#three Accessed 18 Jun 2018
70. Tom Jefferson, Fiona Godlee. Peer Review in Health Sciences, 2nd Edition. Available from: https://www.wiley.com/en-us/Peer+Review+in+Health+Sciences%2C+2nd+Edition-p-9780727916853 Accessed 18 Jun 2018
71. Finelli C, Crispino P, Gioia S, La Sala N, D'amico L, La Grotta M, et al. Retraction: The improvement of large High-Density Lipoprotein (HDL) particle levels, and presumably HDL metabolism, depend on effects of low-carbohydrate diet and weight loss. EXCLI J. 2016 Sep 19;15:570.
72. Finelli C, Crispino P, Gioia S, La Sala N, D'amico L, La Grotta M, et al. The improvement of large High-Density Lipoprotein (HDL) particle levels, and presumably HDL metabolism, depend on effects of low-carbohydrate diet and weight loss. EXCLI J. 2016;15:166–76.
73. Dansinger M. Dear Plagiarist: A Letter to a Peer Reviewer Who Stole and Published Our Manuscript as His Own. Ann Intern Med. 2017 Jan 17;166(2):143.
74. Stigbrand T. Retraction Note to multiple articles in Tumor Biology. Tumor Biol. 2017 Apr 20;1–6.
75. McCook AA. Springer, BMC retracting nearly 60 papers for fake reviews and other issues. Retraction Watch. 2016. Available from: https://retractionwatch.com/2016/11/01/springer-bmc-retracting-nearly-60-papers-for-fake-reviews-and-other-issues/ Accessed: 01/ Jun 2018
76. Cyranoski D. China cracks down on fake peer reviews. Nat News. 2017 Jun 22;546(7659):464.

77. Callaway E. Faked peer reviews prompt 64 retractions. Nat News. Available from: http://www.nature.com.gate2.inist.fr/news/faked-peer-reviews-prompt-64-retractions-1.18202 Accessed 14 Apr 2018
78. McCook AA. A new record: Major publisher retracting more than 100 studies from cancer journal over fake peer reviews. Retraction Watch. 2017. Available from: https://retractionwatch.com/2017/04/20/new-record-major-publisher-retracting-100-studies-cancer-journal-fake-peer-reviews/ Accessed: 01 Mar 2018
79. Mathieu S, Chan A-W, Ravaud P. Use of Trial Register Information during the Peer Review Process. PLOS ONE. 2013 Apr 10;8(4):e59910.
80. Stephenson W. Correlating Persons Instead of Tests. J Pers. 1935 Sep 1;4(1):17–24.
81. FlashQ - Copyright and Disclaimer. Available from: http://www.hackert.biz/flashq/disclaimer/ Accessed 26 Apr 2018
82. Kravitz RL, Franks P, Feldman MD, Gerrity M, Byrne C, Tierney WM. Editorial Peer Reviewers' Recommendations at a General Medical Journal: Are They Reliable and Do Editors Care? PLOS ONE. 2010 Apr 8;5(4):e10072.
83. Henderson M. Problems with peer review. BMJ. 2010 Mar 15;340:c1409.
84. Stahel PF, Moore EE. Peer review for biomedical publications: we can improve the system. BMC Med. 2014 Sep 26;12(1):179.
85. Smith R. Peer review: a flawed process at the heart of science and journals. J R Soc Med. 2006 Apr;99(4):178–82.
86. Baxt WG, Waeckerle JF, Berlin JA, Callaham ML. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. Ann Emerg Med. 1998 Sep;32(3 Pt 1):310–7.
87. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008 Jan 17;358(3):252–60.
88. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ. 2003 May 31;326(7400):1171–3.
89. Rennie D. Editorial peer review: its development and rationale. Available from: http://www.bmj.com/sites/default/files/attachments/resources/2011/07/rennie.pdf Accessed 01 Nov 2017
90. Galipeau J, Moher D, Campbell C, Hendry P, Cameron DW, Palepu A, et al. A systematic review highlights a knowledge gap regarding the effectiveness of health-related training programs in journalology. J Clin Epidemiol. 2015 Mar;68(3):257–65.
91. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. Clin Trials Lond Engl. 2007;4(2):125–39.
92. Cochrane Handbook for Systematic Reviews of Interventions. Available from: http://handbook-5-1.cochrane.org/ Accessed 01 Nov 2017
93. Goodman SN. Manuscript Quality before and after Peer Review and Editing at Annals of Internal Medicine. Ann Intern Med. 1994 Jul 1;121(1):11.
94. 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. Ann Intern Med. 2001 Apr 17;134(8):663–94.
95. van Rooyen S, Black N, Godlee F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. J Clin Epidemiol. 1999 Jul;52(7):625–9.
96. Callaham ML, Baxt WG, Waeckerle JF, Wears RL. Reliability of editors' subjective quality ratings of peer reviews of manuscripts. JAMA. 1998 Jul 15;280(3):229–31.

97. 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 Mar 20;328(7441):673.
98. Callaham ML, Knopp RK, Gallagher EJ. Effect of written feedback by editors on quality of reviews: two randomized trials. JAMA. 2002 Jun 5;287(21):2781–3.
99. Callaham ML, Schriger DL. Effect of structured workshop training on subsequent performance of journal peer reviewers. Ann Emerg Med. 2002 Sep;40(3):323–8.
100. Houry D, Green S, Callaham M. Does mentoring new peer reviewers improve review quality? A randomized trial. BMC Med Educ. 2012 Aug 28;12:83.
101. Arnau C, Cobo E, Ribera JM, Cardellach F, Selva A, Urrutia A. [Effect of statistical review on manuscript quality in Medicina Clínica (Barcelona): a randomized study]. Med Clin (Barc). 2003 Nov 22;121(18):690–4.
102. 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 Mar 28;2(3):e332.
103. Cobo E, Cortés J, Ribera JM, Cardellach F, Selva-O'Callaghan A, Kostov B, 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 Nov 22;343:d6783.
104. 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 Oct;12(5):210–3.
105. 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 Jan 2;318(7175):23–7.
106. 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 Nov 16;341:c5729.
107. 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 Aug;59(8):A4479.
108. Walsh E, Rooney M, Appleby L, Wilkinson G. Open peer review: a randomised controlled trial. Br J Psychiatry J Ment Sci. 2000 Jan;176:47–51.
109. 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 Jul 15;280(3):234–7.
110. Alam M, Kim NA, Havey J, Rademaker A, Ratner D, Tregre B, et al. Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: a randomized multi-rater study. Br J Dermatol. 2011 Sep;165(3):563–7.
111. 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 Jul 15;280(3):240–2.
112. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. A randomized trial. JAMA. 1990 Mar 9;263(10):1371–6.
113. Pitkin RM, Burmeister LF. Identifying manuscript reviewers: randomized comparison of asking first or just sending. JAMA. 2002 Jun 5;287(21):2795–6.
114. Johnston SC, Lowenstein DH, Ferriero DM, Messing RO, Oksenberg JR, Hauser SL. Early editorial manuscript screening versus obligate peer review: a randomized trial. Ann Neurol. 2007 Apr;61(4):A10-12.
115. Neuhauser D, Koran CJ. Calling Medical Care reviewers first. A randomized trial. Med Care. 1989 Jun;27(6):664–6.
116. Rennie D, Knoll E, Flanagin A. The international congress on peer review in biomedical publication. JAMA. 1989 Feb 3;261(5):749.

117. COST | New Frontiers of Peer Review (PEERE). Available from: http://www.cost.eu/COST_Actions/tdp/TD1306 Accessed 28 Nov 2017
118. Glonti K, Cauchi D, Cobo E, Boutron I, Moher D, Hren D. A scoping review protocol on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. BMJ Open. 2017 Oct 22;7(10):e017468.
119. Bohannon J. Who's Afraid of Peer Review? Science. 2013 Oct 4;342(6154):60–5.
120. Walbot V. Are we training pit bulls to review our manuscripts? J Biol. 2009 Mar;8(3):24.
121. 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 Med. 2015 Dec;13(1):158.
122. MacCarthy A, Kirtley S, de Beyer JA, Altman DG, Simera I. Reporting guidelines for oncology research: helping to maximise the impact of your research. Br J Cancer. 2018 Mar 6;118(5):619–28.
123. Munafò M. Metascience: Reproducibility blues [Internet]. Nature. 2017 [cited 2017 Dec 18]. Available from: https://www.nature.com/articles/543619a
124. Hays M, Andrews M, Wilson R, Callender D, O'Malley PG, Douglas K. Reporting quality of randomised controlled trial abstracts among high-impact general medical journals: a review and analysis. BMJ Open. 2016 28;6(7):e011082.
125. Song SY, Kim B, Kim I, Kim S, Kwon M, Han C, et al. Assessing reporting quality of randomized controlled trial abstracts in psychiatry: Adherence to CONSORT for abstracts: A systematic review. PloS One. 2017;12(11):e0187807.
126. Rennie D. Trial registration: a great idea switches from ignored to irresistible. JAMA. 2004 Sep 15;292(11):1359–62.
127. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles. JAMA. 2004 May 26;291(20):2457–65.
128. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Mar 24;340:c869.
129. Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. BMC Med. 2015 Dec;13(1):221.
130. Pocock SJ, Hughes MD, Lee RJ. Statistical Problems in the Reporting of Clinical Trials. N Engl J Med. 1987 Aug 13;317(7):426–32.
131. Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. Lancet Lond Engl. 1990 Jan 20;335(8682):149–53.
132. Andrew E, Anis A, Chalmers T, Cho M, Clarke M, Felson D, et al. A Proposal for Structured Reporting of Randomized Controlled Trials. JAMA. 1994 Dec 28;272(24):1926–31.
133. Call for Comments on a Proposal To Improve Reporting of Clinical Trials in the Biomedical Literature. Ann Intern Med. 1994 Dec 1;121(11):894.
134. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement. JAMA. 1996 Aug 28;276(8):637–9.
135. Chalmers I. Current Controlled Trials : an opportunity to help improve the quality of clinical research. Trials. 2000 Apr;1(1):3.
136. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet

Lond Engl. 2001 Apr 14;357(9263):1191–4.
137. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLOS ONE. 2008 Aug 28;3(8):e3081.
138. Ethgen M, Boutron L, Steg PG, Roy C, Ravaud P. Quality of reporting internal and external validity data from randomized controlled trials evaluating stents for percutaneous coronary intervention. BMC Med Res Methodol. 2009 Apr 9;9:24.
139. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet Lond Engl. 2009 Jul 4;374(9683):86–9.
140. Moher D, Ravaud P. Increasing the evidence base in journalology: creating an international best practice journal research network. BMC Med. 2016 Oct 10;14(1):154.
141. Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. BMJ. 2012 Jun 22;344:e4178.
142. Peters JPM, Hooft L, Grolman W, Stegeman I. Assessment of the quality of reporting of randomised controlled trials in otorhinolaryngologic literature - adherence to the CONSORT statement. PloS One. 2015;10(3):e0122328.
143. Greenwood DC, Freeman JV. How to spot a statistical problem: advice for a non-statistical reviewer. BMC Med. 2015 Dec;13(1):270.
144. Elmore SA. Update on the Manuscript Peer Review Process. Toxicol Pathol. 2017 Dec;45(8):1028–31.
145. Jericho BG, Simpson D, Sullivan GM. Developing Your Expertise as a Peer Reviewer. J Grad Med Educ. 2017 Feb 7;9(2):251–2.
146. Lee CJ, Moher D. Promote scientific integrity via journal peer review data. Science. 2017 Jul 21;357(6348):256–7.
147. Gall T, Ioannidis JPA, Maniadis Z. The credibility crisis in research: Can economics tools help? PLOS Biol. 2017 Apr 26;15(4):e2001846.
148. 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 Oct 1;113(1):651–71.
149. Roman-Belmonte JM, Corte-Rodriguez HD la, Rodriguez-Merchan EC. How blockchain technology can change medicine. Postgrad Med. 2018 May 4;0(0):1–8.
150. COBPeer. Consort-Based Peer-Review Tool. Available from: https://clinicalepidemio.fr/COBPeer_Tool_V2_final/ Accessed 05/ Jul 2018
151. Parsi K, Elster N. Peering into the Future of Peer Review. Am J Bioeth AJOB. 2018 May;18(5):3–4.
152. Kovanis M, Porcher R, Ravaud P, Trinquart L. Complex systems approach to scientific publication and peer-review system: development of an agent-based model calibrated with empirical journal data. Scientometrics. 2016;106:695–715.
153. Shanah D, Sciences aniel has an M in N, Experimental Ms in, transparency TP from U of CA staunch advocate for, Research R in, Evidence he joined BC in 2013 as AP for M. A peerless review? Automating methodological and statistical review. Research in progress blog. 2016. Available from: http://blogs.biomedcentral.com/bmcblog/2016/05/23/peerless-review-automating-methodological-statistical-review/ Accessed 30 Jun 2018
154. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc JAMIA. 2016 Jan;23(1):193–201.
155. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, et al. Risk of

bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ. 2009 Oct 19;339:b4012.
156. ijmarshall (Iain Marshall). GitHub. Available from: https://github.com/ijmarshall Accessed 28 Jun 2018