



---

# Improving peer review via controlled experiments

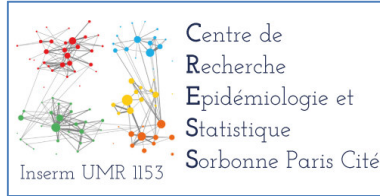
Isabelle Boutron

METHODS team

Research Centre of Epidemiology Biostatistics Sorbonne Paris Cité  
Paris Descartes University



# Background



The focus of the French  
EQUATOR centre is on  
increasing/improving/enhancing  
the value of biomedical  
research

## METHODS Team

Methods of therapeutic evaluation of  
chronic diseases



- Clinical epidemiologist
- Focus on **clinical research (randomized controlled trials)**
- Focus on **Research on Research and particularly interventional research on research**
  - Develop interventions to improve research and use high level evidence study designs or modelling to evaluate these interventions

# Acknowledgements

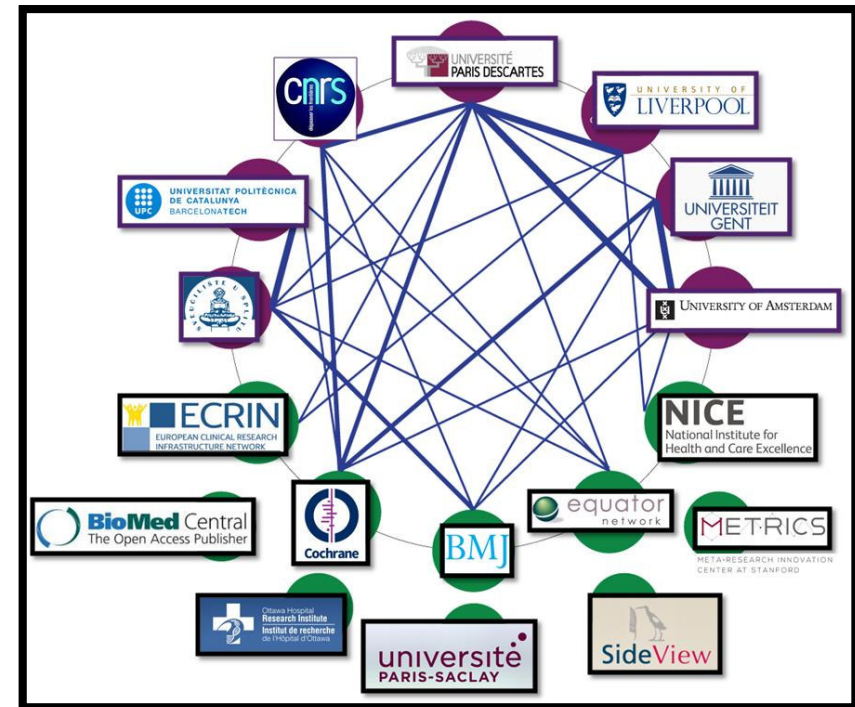
## 7 European Universities and 10 International Partners

### MiRoR project

Joint doctoral training program, to train **15 PhD students in Methods in Research on Research** in the field of clinical research (<http://mirror-ejd.eu>) funded by Marie Skłodowska-Curie Actions.

### Projects

- Planning
- Conduct
- Reporting
- Peer review



# The peer review system

---

- Central to the scientific community
- 2 major goals
  - Gatekeeper of the scientific publications
  - To improve the quality of manuscripts
- A system relying mainly on work performed voluntarily by academic researchers



---

**Is the peer-review system achieving its goal?**

# Testing for the Presence of Positive-Outcome Bias in Peer Review

*Arch Intern Med. 2010;170(21):1934-1939*

## *A Randomized Controlled Trial*

*Gwendolyn B. Emerson, MD; Winston J. Warme, MD; Fredric M. Wolf, PhD;  
James D. Heckman, MD; Richard A. Brand, MD; Seth S. Leopold, MD*

Two versions of a well-designed randomized controlled trial that differed only in the direction of the finding of the principal study end point

238 reviewers at 2 journals (assigned at random)

- were **more likely to recommend the positive** version of the test manuscript for publication than the no-difference version (97% vs 80%,  $P < 0.001$ )
- **detected more errors** in the no-difference version than in the positive version (mean 0.85 vs 0.41,  $P < 0.001$ )
- **awarded higher methods scores** to the positive manuscript than to the (identical) no-difference manuscript (8.24 vs 7.53,  $P = 0.005$ )

September 27, 2016

## Single-blind vs Double-blind Peer Review in the Setting of Author Prestige

Kanu Okike, MD, MPH<sup>1</sup>; Kevin T. Hug, MD<sup>2</sup>; Mininder S. Kocher, MD, MPH<sup>3</sup>; [et al](#)

- 119 reviewers were randomized to assess a fabricated manuscript with the prestigious authors' names and institutions masked or visible
  - Reviewers were **more likely to recommend** acceptance when the prestigious authors' names and institutions were visible than when they were redacted
    - 87% vs 68%;
    - RR, 1.28 [95% CI, 1.06-1.39], P = .02
  - They gave **higher ratings** for the methods.

# Transparency– Impact of the peer review process

---



**Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study**

- Comparison of the submitted and accepted versions of manuscripts of RCTs published in open access journals (n=93)
- Changes were limited
  - A median of **11%** (range 1-60%) **words deleted** and **20% added** (range 2-88%).

# Impact of the peer review process on Transparency and completeness of reporting

- 93 RCTs published
- First submission / Peer-reviewers' comments/ Published article

Adequate reporting	Submitted	Published
Sequence generation	47%	59%
Allocation concealment	34%	44%
Blinding	33%	45%
Primary outcome	51%	51%
Results for the primary outcome	35%	35%

# Transparency of published reports (peer-reviewed) vs posting in trial registries (no peer-review)



## Timing and Completeness of Trial Results Posted at ClinicalTrials.gov and Published in Journals

Sample of randomized controlled trials of drugs with results both posted and published

Completeness of reporting	ClinicalTrials.gov N=202	Published article N=202	P-value
Flow of participants	64%	48%	<0.001
Efficacy results	79%	69%	0.02
Adverse events	73%	45%	<0.001
Serious adverse events	99%	63%	<0.001

# Detection of selective reporting of outcomes

OPEN ACCESS Freely available online

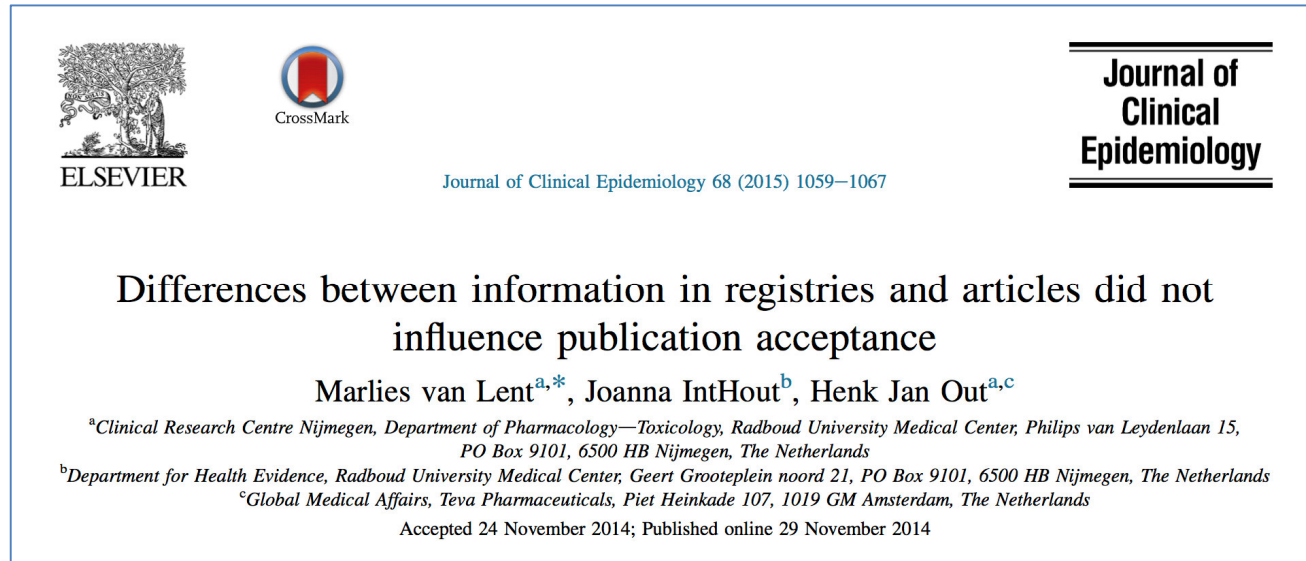


## Use of Trial Register Information during the Peer Review Process

Sylvain Mathieu<sup>1,2</sup>, An-Wen Chan<sup>3</sup>, Philippe Ravaud<sup>1\*</sup>

- Survey of 676 authors and reviewers who had reviewed at least 1 article reporting a clinical trial in the past 2 years
- **34% examined information registered on a trial registry.**

# Detection of selective reporting of outcomes

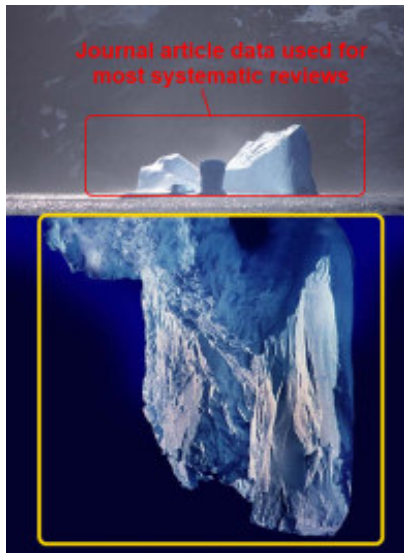


- 226 manuscripts of drug trials submitted to 8 journals
- (29% had changes in the primary outcome
- The pattern of editorial decisions was not statistically significantly different for manuscripts with or without changed primary outcomes ( $P = 0.418$ ).



# Can peer reviewers reduce spin ?

**Reporting bias**



**Data beautification  
Spin**



- Spin: A way of reporting to convince the reader that the beneficial effect of the experimental treatment (efficacy, safety) is higher than shown by the results
- Spin is frequent in published reports
  - 50% of abstract conclusions of RCTs have spin

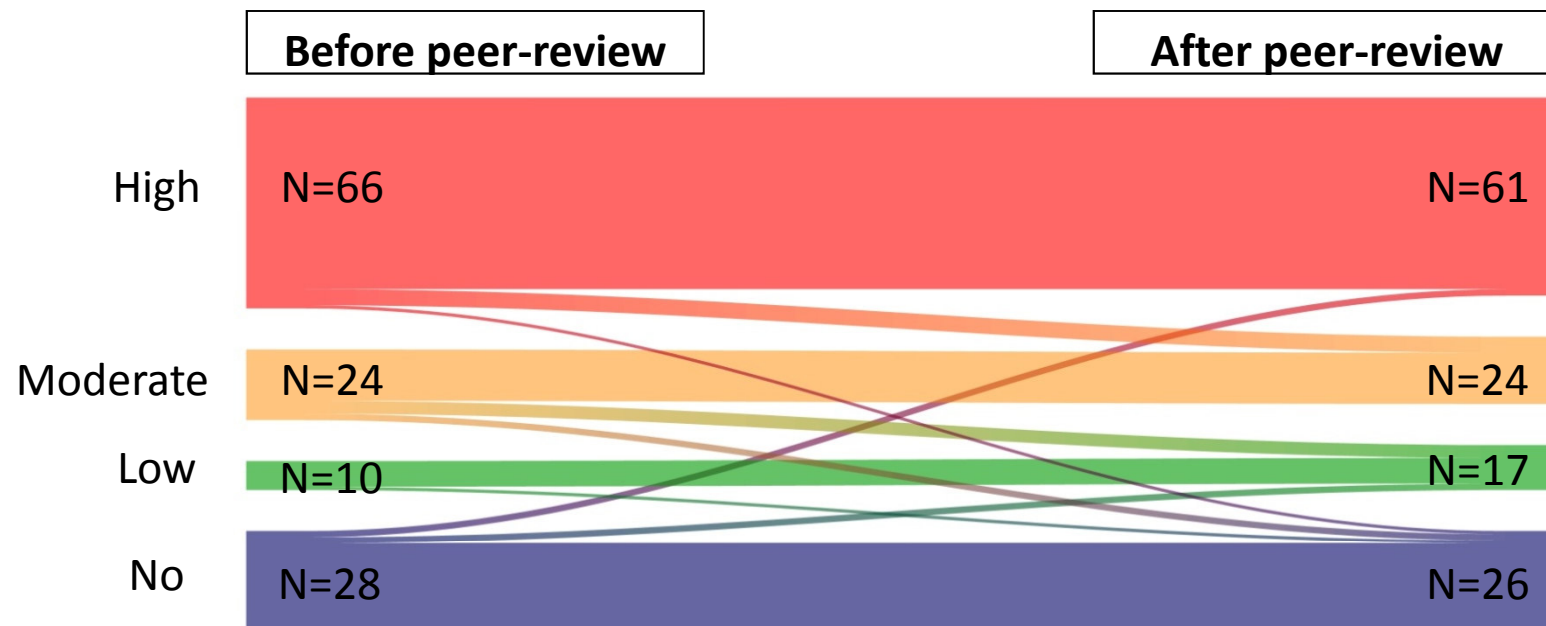
# Can peer reviewers reduce spin ?

---

- Assessment of the manuscript submitted, peer reviewers comments, and final manuscript of non-randomized studies assessing a therapeutic intervention published in open access journals (n=128).
- 55% of submitted manuscripts, peer reviewers identified at least one example of spin
- Of the spin identified by peer reviewers
  - 67% were completely deleted,
  - 16% partially deleted
  - 17% not removed in the final published article.
- For 15%, peer reviewers requested adding some spin

# Level of spin before and after peer review in the abstract conclusion

---



**76% Peer reviewers failed to identify spin in abstract conclusions**

---

## **Controlled experiments to improve the peer-review process**

# Interventions to improve transparency

## CONSORT Statement, recommendations for reporting RCTs

## Extensions of CONSORT

**Annals of Internal Medicine** | **ACADEMIA AND CLINIC**

### CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials

Kenneth F. Schulz, PhD, MBE; Douglas G. Altman, DSc

The CONSORT (Consolidated Standards of Reporting Trials) Statement is a widely used tool to improve the reporting of randomized trials. Schulz and colleagues describe the CONSORT 2010 Statement, which updates the reporting guidelines to reflect new methodological evidence and accumulating experience.

**Checklist** *Now in order to encourage dissemination, the CONSORT 2010 Statement, the checklist is free to use on www.consort-statement.org and will also be published in Lancet, Obstetrics & Gynaecology, PLoS Medicine, Medicine, Journal of Clinical Epidemiology, Statistics, and Trials. The authors jointly hold the copyright. For details on further use, see the CONSORT website: www.consort-statement.org.*

Randomized, controlled trials, when appropriately designed, conducted, and reported, represent the gold standard in evaluating health care interventions. However, randomized trials can yield biased results if methodological errors (1). To assess a trial accurately, a published report must provide complete, clear, and transparent information on its methodology and findings. Unfortunately, attempted assessments frequently find that authors of many trial reports neglect to provide complete descriptions of their critical information. This lack of adequate reporting fueled the need for the original CONSORT (Consolidated Standards of Reporting Trials) Statement in 1996 (2) and its 5 years later (3–6). While those statements improved reporting quality for some randomized, controlled trials, many trial reports still remain inadequate (2). More, new methodological evidence and additional data have accumulated since the last revision in 2007. Consequently, we organized a CONSORT Group to update the 2007 statement (3–6). We introduce the result of that process, CONSORT 2010.

**INTENT OF CONSORT 2010**

The CONSORT 2010 Statement is the updating of the 25-item checklist in the Table and diagram (Figures). It provides guidance for reporting randomized, controlled trials but focuses on the main design type—individually randomized, 2-group, parallel trials. Other trial designs, such as cluster trials and noninferiority trials, require varying additional information. CONSORT extension designs (11, 12), and other CONSORT products found through the CONSORT Web site (www.consort.org). Along with the CONSORT

Section/Topic	Item Number	Checklist Item	Reported on Page Number
Introduction	1a	Identification as a randomized trial in the title	1a
Background and objectives	1b	Structured summary of trial design, methods, results, and conclusions; the specific questions, see CONSORT for details (2, 11)	1b
Methods	2a	Isolate background and separation of rationale	2a
Study design	2b	Specify objectives or hypothesis	2b
Participants	3a	Description of trial design (such as parallel, factorial, including allocation ratio) important changes to methods after trial commencement (such as eligibility criteria, etc.)	3a
Interventions	3b	Eligibility criteria for participants	3b
Outcomes	3c	Why were individuals excluded from the study? Were they actually allocated?	3c
Sample size	3d	How many individuals were included? Including how and when they were assessed	3d
Randomization	4a	How many individuals were included? Including how and when they were assessed	4a
Blinding	4b	How many individuals were included? Including how and when they were assessed	4b
Statistical methods	4c	How many individuals were included? Including how and when they were assessed	4c
Results	4d	How many individuals were included? Including how and when they were assessed	4d
Discussion	4e	How many individuals were included? Including how and when they were assessed	4e
Conclusions	4f	How many individuals were included? Including how and when they were assessed	4f

### RESEARCH METHODS & REPORTING

#### CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials

David Moher, Sally Hopewell, Kenneth F Schulz, Vitoria M Altman, Peter G Gøtzsche, P J Vandenbroucke, Diana Elbourne, Matthias Egger, Douglas G Altman

**ABSTRACT**

Randomized trials are the gold standard for evaluating the quality of reporting of randomized controlled trials (RCTs) and optimal. Without transparent reporting, randomization cannot be used to evaluate the quality of evidence for specific interventions. However, methodological errors in reporting of RCTs have been identified, and these errors have led to biased estimates of treatment effects. Such errors are most likely to occur in the reporting of RCTs, which are considered the gold standard for evaluating interventions because of their ability to minimize bias and error.

A group of international experts developed the CONSORT 2010 Statement to provide guidance for reporting RCTs. The CONSORT 2010 Statement is a 25-item checklist and explanation and elaboration of the checklist. The checklist is intended to be used by authors, reviewers, editors, and readers of RCTs. The explanation and elaboration provides detailed guidance for each item in the checklist. The CONSORT 2010 Statement is available at [www.consort-statement.org](http://www.consort-statement.org).

**BMJ**

RESEARCH METHODS & REPORTING

### CONSORT 2010 statement extension to cluster randomised trials

Reporting of Noninferiority and Equivalence Randomized Trials

CONSORT Statement for Reporting Randomized Controlled Trials in Journal and Conference Abstracts: Explanation and Elaboration

CONSORT Statement for Reporting Methods of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts

- CONSORT: 25-item checklist + E&E paper
- Requested in instructions to authors of several journals
- Authors are requested to submit the relevant checklist
- Adherence of authors to these guidelines remains low

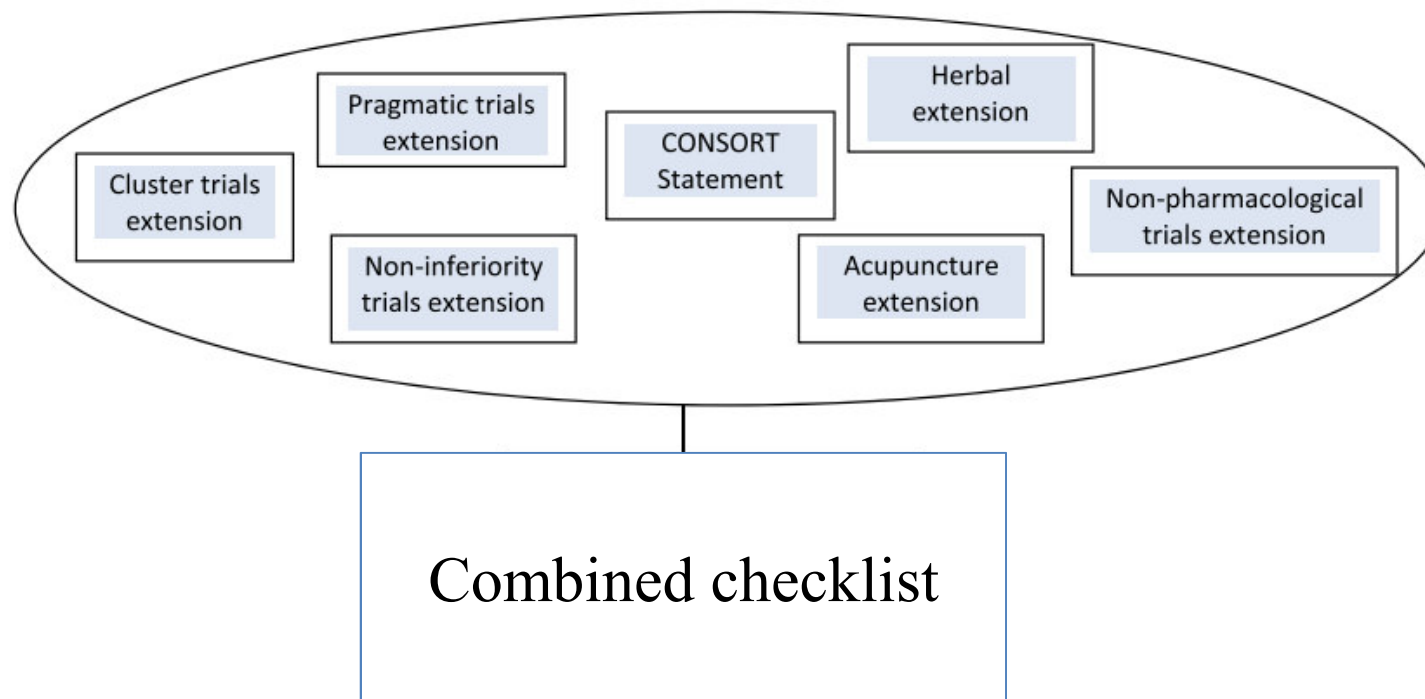
RESEARCH ARTICLE

Open Access



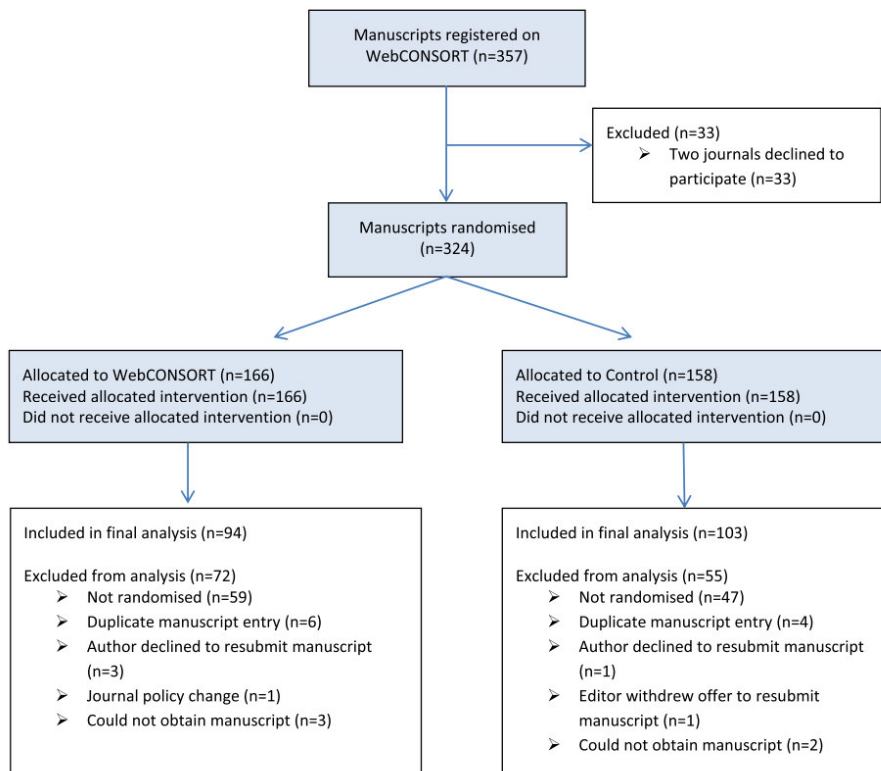
# Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial

Sally Hopewell<sup>1,2,3\*</sup>, Isabelle Boutron<sup>3,4</sup>, Douglas G. Altman<sup>2</sup>, Ginny Barbour<sup>5</sup>, David Moher<sup>6</sup>, Victor Montori<sup>7</sup>, David Schriger<sup>8</sup>, Jonathan Cook<sup>2</sup>, Stephen Gerry<sup>2</sup>, Omar Omar<sup>2</sup>, Peter Dutton<sup>2</sup>, Corran Roberts<sup>2</sup>, Eleni Frangou<sup>2</sup>, Lei Clifton<sup>2</sup>, Virginia Chiochia<sup>2</sup>, Ines Rombach<sup>2</sup>, Karolina Wartolowska<sup>2</sup>, and Philippe Ravaud<sup>3,4</sup>

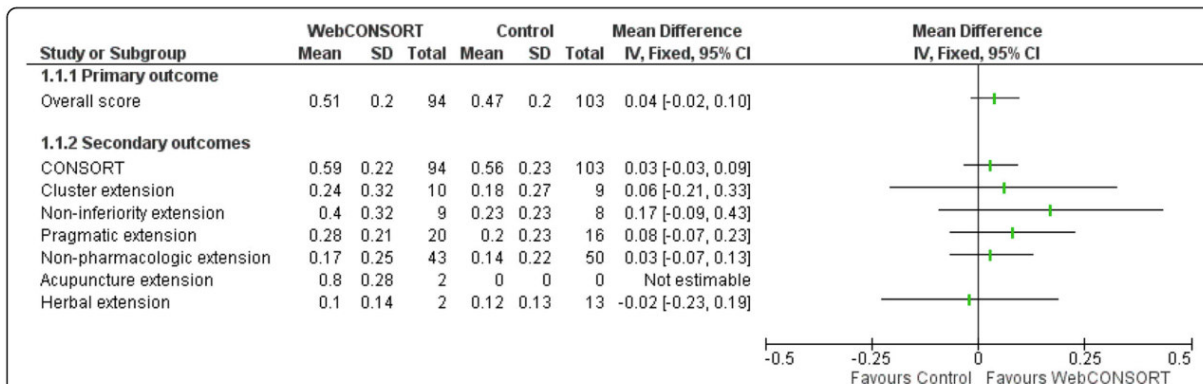




- 46 journals actively recruited into the trial
- 324 manuscripts were randomised



**Fig. 2** Flow of manuscripts registered on the WebCONSORT study website



**Fig. 3** Comparison of overall mean score between WebCONSORT and Control interventions ( $n = 197$  manuscripts)

**1/3 manuscripts selected for RCT by the editorial staff were not randomised**

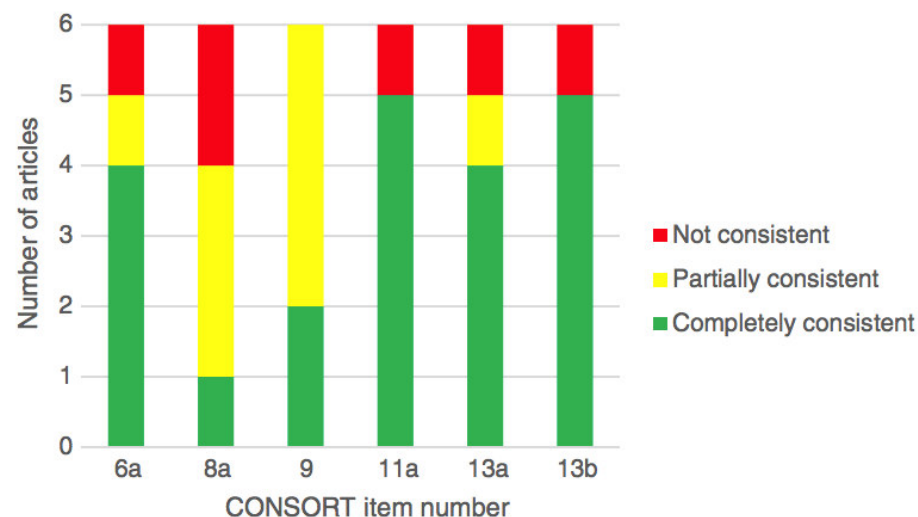
COMMENTARY

Open Access



# Are CONSORT checklists submitted by authors adequately reflecting what information is actually reported in published papers?

David Blanco<sup>1,3\*</sup> , Alice M. Biggane<sup>2,3</sup>, Erik Cobo<sup>1</sup> and MiRoR network



**Fig. 1** Reporting inconsistencies found for the six papers that used the appropriate CONSORT checklist



## Format

- Checklist of items: not understood by most authors
- Elaboration & Explanation manuscript: >30 pages (never read by authors)



## Format

- Template with the checklist tailored and expanded with details on what should be reported for each item

## Writing Aid Tool

### Expanded, combined and tailored checklist

## Checklist

## E & E

**Elaboration Explanation manuscript**

Evidence suggests that patient outcome can be associated with hospital and care provider volume (15, 55-57). A systematic review of 15 trials (15) found that 71% observed a positive association between hospital volume and outcomes and 70% observed an association between care provider volume and outcomes. Differential expertise of care providers in each treatment group on the treatment effect estimate (54). Furthermore, a nonpharmaceutical treatment might be found to be safe and effective in an RCT performed in high-volume centers by high-volume care providers, but could have different results in low-volume centers. For example, the Asymptomatic Carotid Atherosclerosis Study investigators included 80% of all possible care providers, selecting only those with good safety records. This resulted in a prospective mortality rate that was 5 times lower than in other trials with less stringent selection criteria (58-61). In most nonpharmaceutical trials, care provider expertise and center volume of care will influence the treatment effect (11, 55, 57, 62-75).

Reporting of eligibility criteria for care providers and centers in nonpharmaceutical trials is often poor. One study of surgical reports found that the setting and the center volume of activity was reported in only 7% and 3% of articles, respectively (76). Selection criteria were reported for care providers in 41% of the articles, and the number of care providers performing the intervention was reported in 32% (76). A careful description of care providers involved in the trial, as well as details of the centers in which participants were treated, helps readers appraise the risk for bias and the applicability of the results. Selection criteria for centers typically relate to center volume for the procedure under investigation or similar procedures. Eligibility of care providers might include professional qualifications, years in practice, medical or intervention performed, skill as assessed by level of complication when performing the intervention, and specific training before trial initiation. Eligibility criteria should be justified, because they will influence the applicability of the trial results (58, 73-74).

## Item 6 a

*“Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed”*

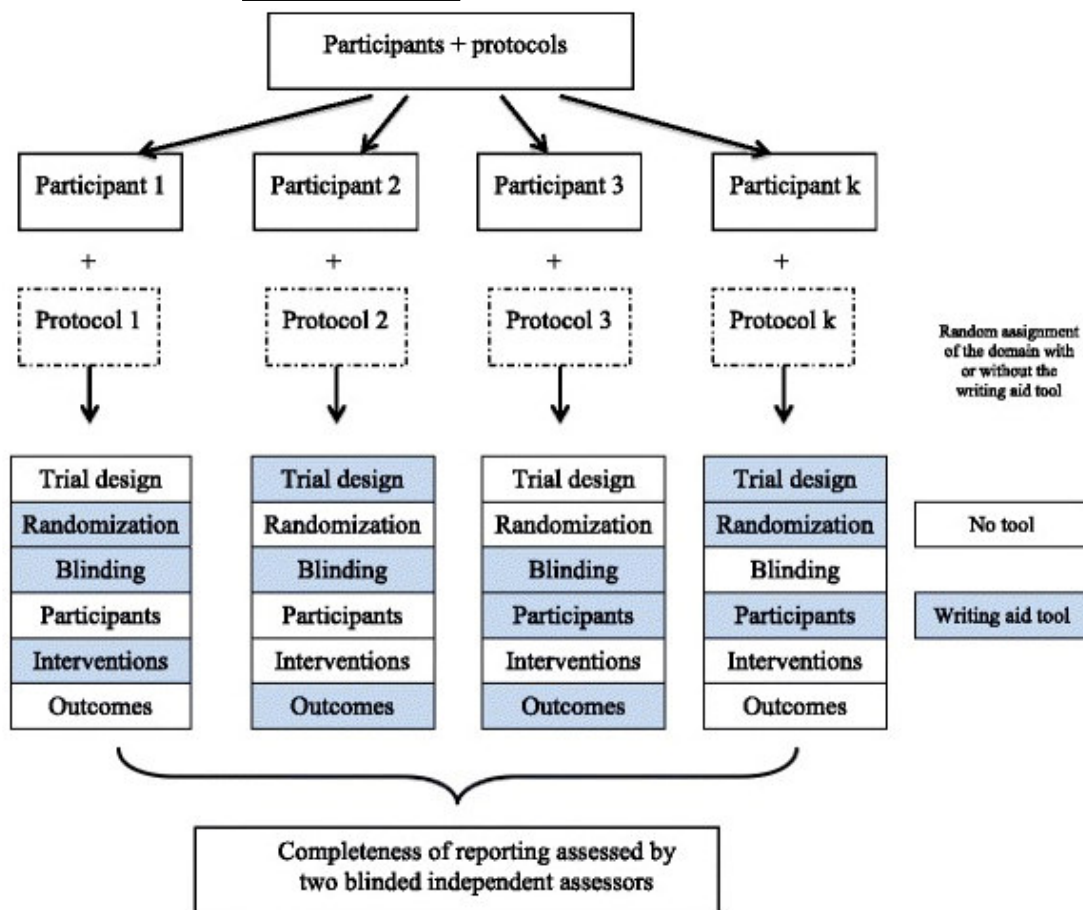


## Please report

- Which outcome(s) is the primary outcome(s) (i.e. pre-specified outcome considered to be of greatest importance).

## For each primary and secondary outcome, report

- The variable of interest (eg, pain)
- How the outcome was assessed (eg, VAS)
- The analysis metric (eg, change from baseline)
- Time point of interest for analysis (eg, 3 months)
- ...



- Design: “**Split-manuscript randomised controlled trial**” with blinded outcome assessment
- 41 students (masters and PhD)
- Session: 4 hours

**Completeness of reporting (0-10) writing aid tool vs none**

Mean Difference = 2.1 [1.5;2.7];  
p <0.001

**Completeness of reporting (0- 10) writing aid tool vs NEJM/JCO**

Mean Difference = 1.7 [1.1; 2.4];  
p <0.001

**test**

Invite a reviewerExport paperExport paper with Consort Items

Title

Introduction

Methods

**Trial design**

Participants

Interventions

Outcomes

Sample size

Randomisation

Blinding

Statistical methods

Results

Save changes

**Description of trial design (such as parallel, factorial) including allocation ratio**

Please report:

- The type of trial design (parallel group)
- The conceptual framework (superiority, non-inferiority, equivalence, other)
- The allocation ratio
- Any other pertinent information (for drug development (phase 1,2,3), other)

**Important changes to methods after trial commencement (such as eligibility criteria), with reasons**

**Example**

"This was a multicenter, stratified (6 to 11 years and 12 to 17 years of age, with imbalanced randomisation [2:1]), double-blind, placebo-controlled, parallel-group study conducted in the United States (41 sites)." (Blumer et al., Pediatrics 2009)

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



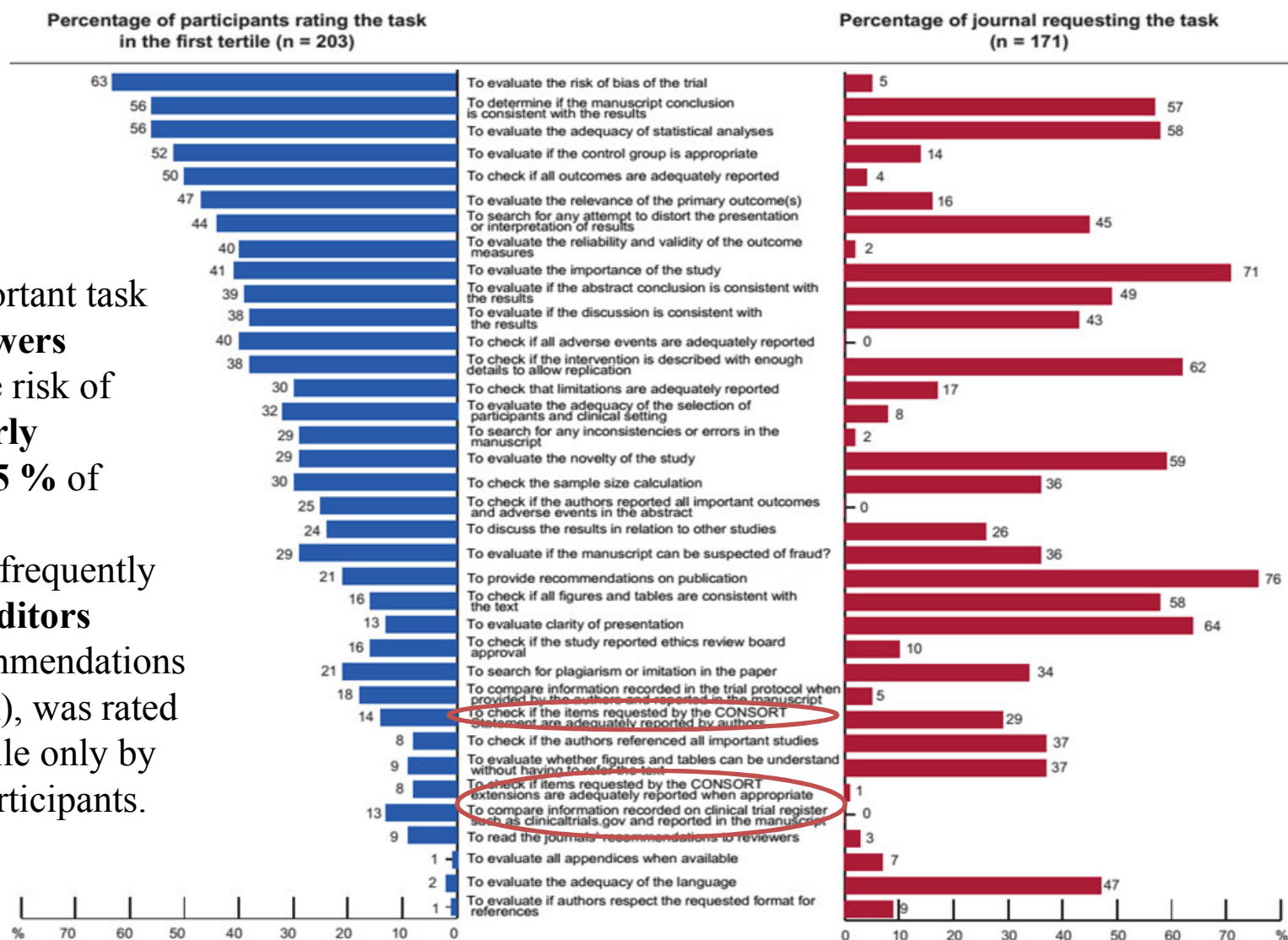
Anthony Chauvin<sup>2,3†</sup>, Philippe Ravaud<sup>1,2,3</sup>, Gabriel Baron<sup>1,3</sup>, Caroline Barnes<sup>2,3</sup> and Isabelle Boutron<sup>1,2,3\*†</sup>

- The tasks expected from a peer-reviewer are not realistic
  - More than 200 different tasks identified
  - The tasks involved different level of expertise and different backgrounds
    - Statistical and methodologic expertise
    - Content expertise (novelty, relevance, external validity etc)
    - None
      - Verification (adherence to guidelines, consistency with registries)
      - Formatting

# Multiple tasks asked to reviewers (important tasks for reviewers are not congruent with important tasks for editors)

*Chauvin, BMC Medicine, 2016*

- The most important task for **peer reviewers** (evaluating the risk of bias) was **clearly requested by 5 %** of editors.
- The task most frequently requested by **editors** (provide recommendations for publication), was rated in the first tertile only by **21 %** of all participants.



## **A 2-step peer-review system to improve completeness of reporting. Theoretical background of the intervention**

- Peer-reviewers consider that assessing adherence to reporting guidelines and switch in outcomes is not their remit.
- Detection of inadequate reporting does not involve high level of expertise and could be performed by early career researchers who would gain expertise in peer-review.
- Online tool and a training module could early career researchers detected misreported items and switched outcomes when evaluating a report of a RCT.



# Development of the intervention - COBPeer

COBPeer A Consort-Based Peer-Review Tool

**The ARTIST (osteoARThritis Intervention Standardized) study:**

**A pragmatic randomised controlled trial comparing standardized consultation to usual care for patients with knee osteoarthritis under primary care**

P. Ravaud (MD, PhD)<sup>1</sup>, R-M Flipo (MD)<sup>2</sup>, I. Boutron (MD, PhD)<sup>1</sup>, C. Roy (MsC)<sup>1</sup>, A. Mahmoudi (MD)<sup>3</sup>, B. Giraudeau<sup>3</sup> (PhD), T. Pham (MD)<sup>4</sup>

<sup>1</sup> INSERM, U738, Paris, France ; Université Paris 7 Denis Diderot, UFR de Médecine, Paris, France  
; AP-HP, Hôpital Bichat, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Paris  
, France

**OUTCOMES**  
Completely defined pre-specified primary outcome measures, including how and when they were assessed

• Was the primary outcome(s) clearly identify ? ☐ Yes ☐ No

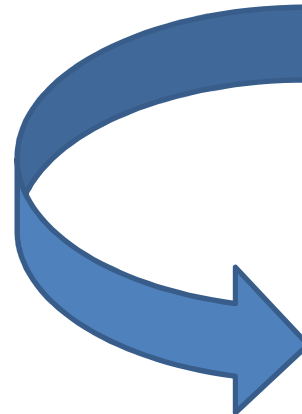
**RANDOMIZATION / SEQUENCE GENERATION**  
Method used to generate the random allocation sequence

Did the author report:

• The method of sequence generation (e.g., a random number table or computerized random number generator, or other) ☐ Yes ☐ No

**ALLOCATION CONCEALMENT**  
Mechanism used to implement the random allocation sequence (e.g., sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

Did the author report:



## Automatically generated peer-review report

**Peer-Review report**

Dear Authors,

Please note that the completeness of reports of your manuscript has been assessed and some essential information is missing. We have summarized below the list of missing information that should be reported before sending your manuscript to our external peer review.

Should you need any further information, please do not hesitate to contact us.

Sincerely,

**Primary Outcome(s)**

- Please could you clearly report which outcome is the primary outcome (i.e., the pre-specified outcome considered to be of greatest importance and usually used in the sample size calculation)  
- For the primary outcome please clearly report the following information

- The **variable of interest** (e.g., pain, Beck Depression Inventory score, all-cause mortality)
- **How** the variable of interest was measured (e.g., VAS, Beck depression Inventory Score)
- 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 point** of interest for analysis (e.g., one hour, 3 months)
- **How** the outcome was assessed (e.g., EVA)

**Participant flow**

- Please report a **flowchart**

In the flow diagram or in the text, please report:

- Number of participants **lost to follow-up with reasons in each group**
- Number of participants **analyzed** for the primary outcome in each group

**Outcomes and estimation**

For the primary outcome, please report:

- The **precision** for difference between groups (e.g., 95% CI)

**Harms.**

**Please report:**

- For each group, the **number of participant withdrawals** due to harm
- Results in each group for each **adverse event type** (mean (SD) or number of event/n)

**Registration**

- If you registered your protocol, please reported the registration number. If you did not register your protocol, please give your reasons.

# Development of the intervention – COBPeer training programme

## COBPeer training module

Item 6a. Completely defined pre-specified primary outcome measures, including how and when they were assessed

"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 ? <input type="radio"/> <input type="radio"/>

Outcomes	Item 6a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed
Yes	

→ Congratulations! Your answers are correct.

"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."

Explanation

Authors reported numbers data collected, but they did not indicate which was the primary outcome. So the primary outcome reported by authors is unclear.

- 1/ The variable of interest: NOT REPORTED BY AUTHORS
- 2/ How the outcome(s) were assessed: NOT REPORTED BY AUTHORS
- 3/ The format of the primary outcome data: NOT REPORTED BY AUTHORS
- 4/ The summary measure for each study group: NOT REPORTED BY AUTHORS
- 5/ The time frame to measure primary outcome: NOT REPORTED BY AUTHORS
- 6/ Who assessed the outcome: NOT REPORTED BY AUTHORS



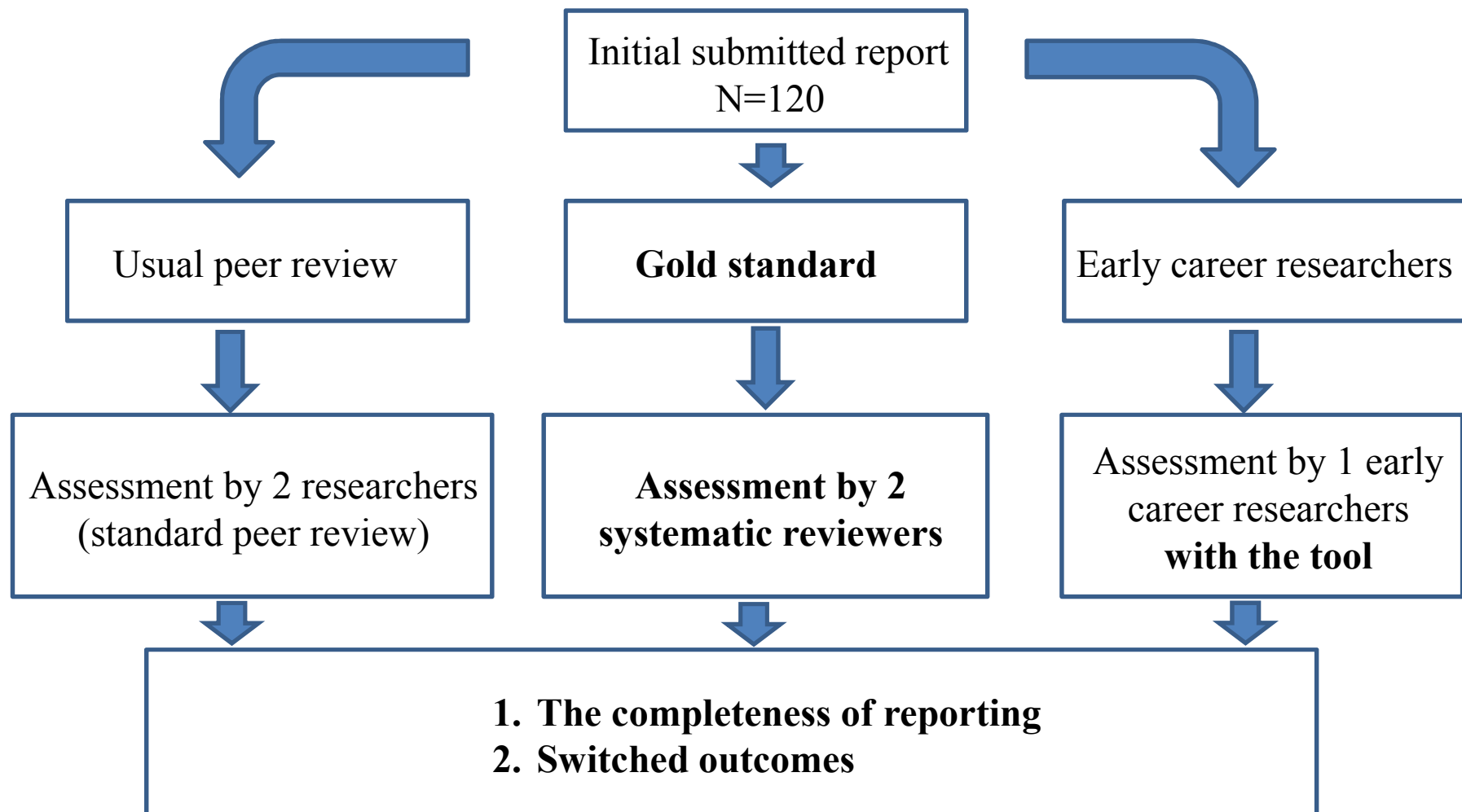
# Assessment of the intervention

---

- **Objective**
  - To compare the performance of early career peer reviewers who use the peer reviewer tool (COBPeer) with usual peer reviewers in identifying incomplete reporting and switched outcomes in reports of RCTs.
- **Study design**
  - **Cross-sectional study** comparing the accuracy of early career peer reviewers using COBPeer 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
  - **Gold standard:** assessment of systematic reviewers

# Assessment of the intervention

---



# Several interventions are implemented or proposed to improve the system

## Elsevier initiative leads to faster revision and review times

Button et al. *BMC Psychology* (2016) 4:59  
DOI 10.1186/s40359-016-0167-7

BMC Psychology

EDITORIAL Open Access

Preventing the ends from justifying the means: withholding results to address publication bias in peer-review

Katherine S. Button<sup>1\*</sup>, Liz Bai<sup>2</sup>, Anna Clark<sup>2</sup> and Tim Shipley<sup>2</sup>

WELCOME TO PUBLONS ACADEMY  
Become a master of peer review

SIGN UP

START BUILDING YOUR PROFILE AS AN EXPERT IN YOUR FIELD  
CONNECT WITH TOP JOURNAL EDITORS  
WORK WITH YOUR SUPERVISOR TO COMPLETE YOUR FIRST REAL REVIEWS

nature COMMUNICATIONS

EDITORIAL OPEN

**Towards an unbiased view of science**

Authors at *Nature Communications* now have the option to choose double-blind peer review

**N**ature Communications is an open journal and our articles can be read online by all. Our aim is to publish high-quality research across all of the natural sciences, reporting discoveries that are important to specialists within their respective research fields. We welcome scientific submissions from anyone, and we aim to select the papers that we publish based on the significance of the science presented without any bias towards an author's ethnicity, nationality, gender, number of prior publications or any other factor.

For those manuscripts passing our editorial screening process, peer reviewers are involved in assessing papers

During double-blind peer review the identity of authors is known to the editors, and some might call for a 'triple-blind' system to eliminate the potential for editorial bias. This would not only be difficult to implement, it would also come at the cost of preventing direct discussions between editors and the authors they serve. However, most of our submissions are assessed by more than one editor, and decisions after double-blind peer review are, of course, made on the basis of the reports of reviewers who are not aware of the authors' identities. A checklist that might be useful to ensure that author identity is not readily apparent to reviewers can be found at <http://www.nature.com/authors/double-blind-checklist.pdf>.

Automated **Statistical Support** for Journals and Authors

"...the majority of statistical analyses are performed by people with an inadequate understanding of statistical methods. They are then peer reviewed by people who are generally no more knowledgeable"

— Douglas Altman

stat reviewer

## Could Robots Handle Peer Review?

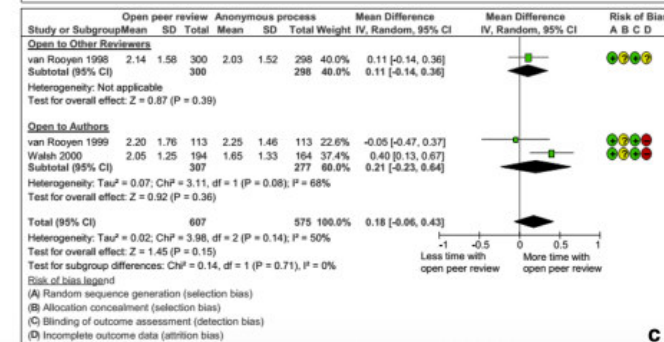
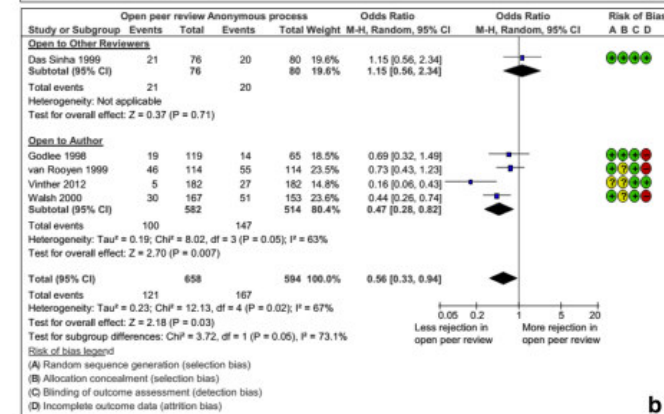
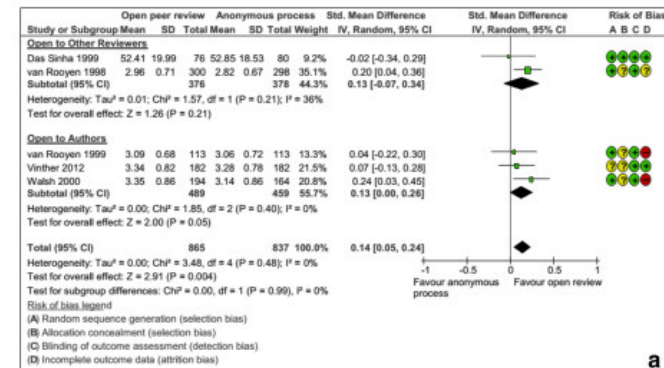
Technologist argues that artificial intelligence could make publishing decisions in milliseconds.

# What is the evidence?

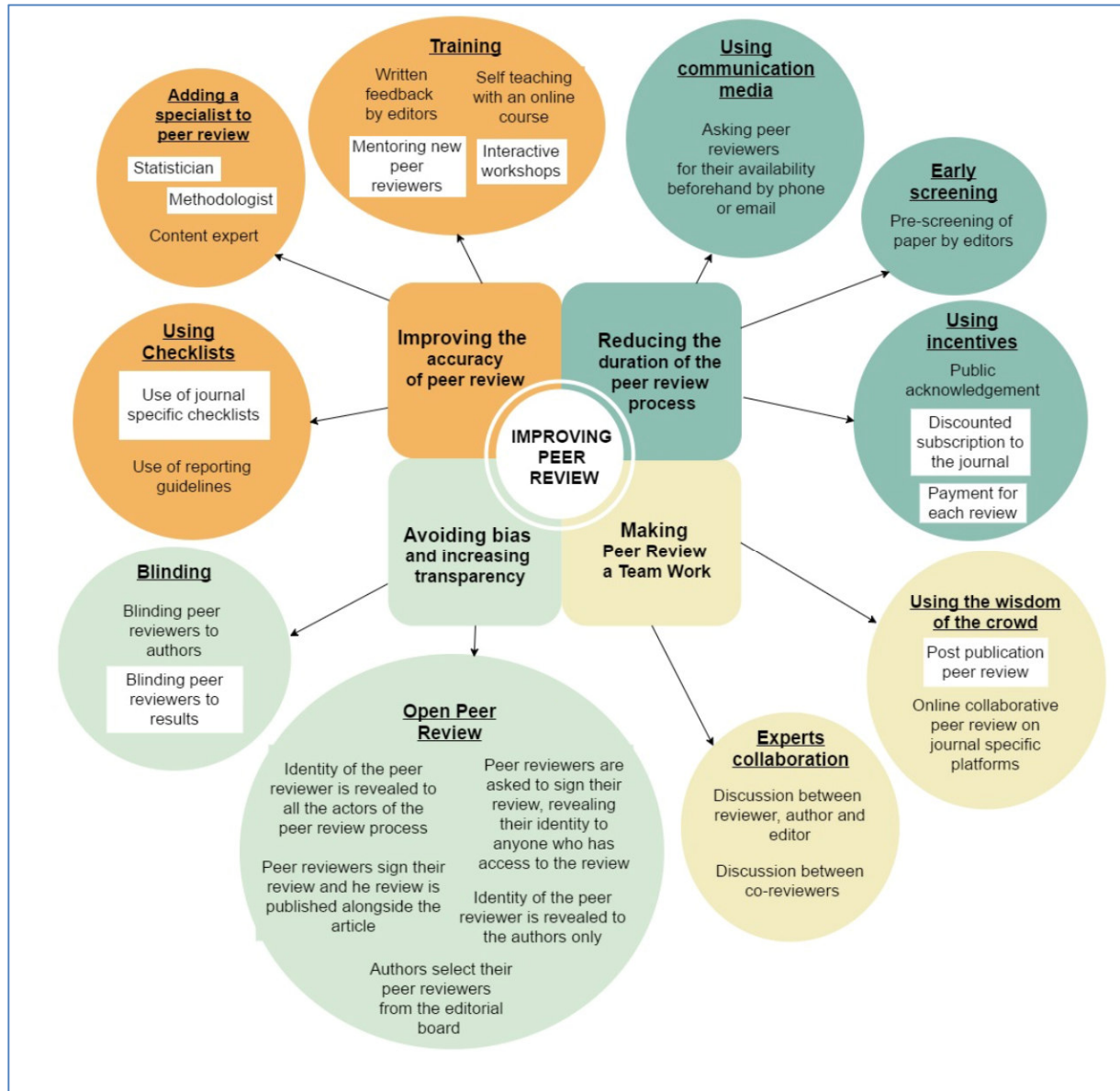
## A Systematic Review and Meta-analysis of Interventions to Improve the Quality of Peer Review

- Only 22 reports of RCTs
- Only 7 were published over the past 10 years
- Interventions assessed
  - Blinding
  - Open peer review
  - Training
  - Use of checklist
  - Adding experts
- Most are performed in one single journal
- Low methodologic quality

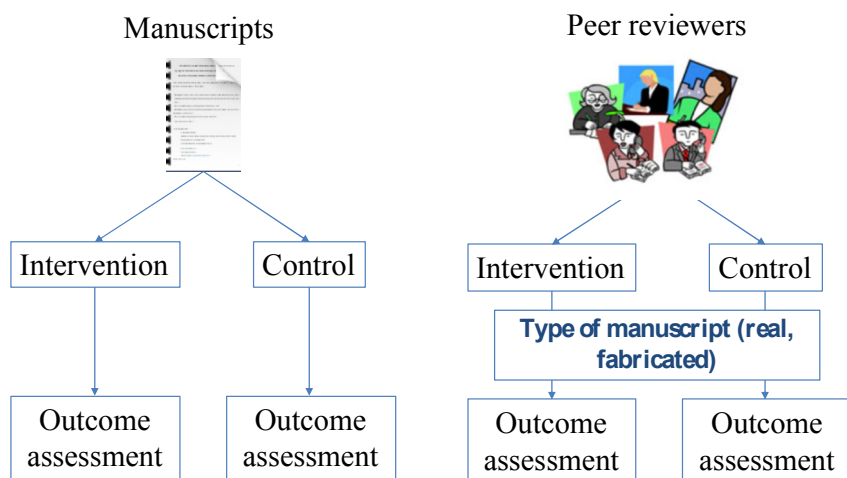
Bruce, BMC Med, 2016



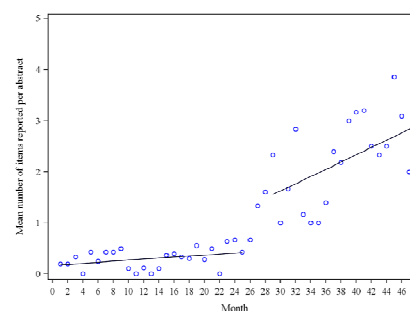
# What should be the study design?



# What should be the study design?



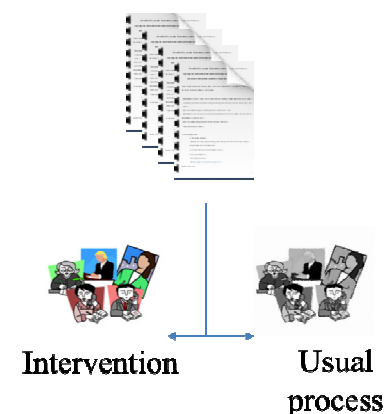
## Time series analysis



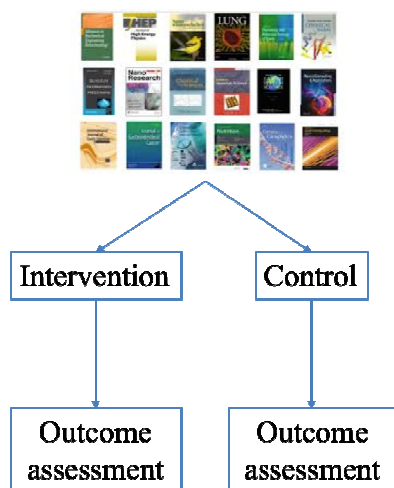
Active implementation  
(Lancet, Ann Intern Med)

Level change:  $p = 0.0035$   
Trend change:  $p = 0.0154$

## Paired study design



## Cluster RCTs



## Stepped-wedge RCT

STEPS (Cluster or Group of Clusters)	Time					
	1	2	3	4	5	6
1	Exposed	Exposed	Exposed	Exposed	Exposed	Exposed
2	Exposed	Exposed	Exposed	Exposed	Exposed	Exposed
3	Exposed	Exposed	Exposed	Exposed	Exposed	Exposed
4	Exposed	Exposed	Exposed	Exposed	Exposed	Exposed
5	Exposed	Exposed	Exposed	Exposed	Exposed	Exposed

Legend:  Exposed to intervention  Unexposed to intervention

- Type of manuscript (real, fabricated)
- Setting single journal, several journals, several publisher)



# What should be the study design?

Fixed for each intervention

Varied within a same intervention

Does the <u>use of reporting guidelines</u> by the peer reviewer improve the quality of the final manuscript, compared to the usual process?	
Intervention	Peer reviewers are asked to fill in a checklist based on guidelines (such as CONSORT or STARD, depending on the nature of their manuscript) in addition to their usual review. The checklist is then sent to the authors so they can revise their manuscript.
Comparator	Peer reviewers are not asked to fill in a guidelines checklist. They follow the usual process of peer review.
Main outcome measure	Quality of the revised manuscript updated by the authors <ul style="list-style-type: none"><li>Measured with a manuscript quality assessment tool: a 5 point Likert scale from 1 (low) to 5 (high) , with 34 items regarding the originality of the paper, the strengths and weaknesses of the method, the presentation, the constructiveness of comments, the substantiation of comments and the interpretation of results</li><li>By a blinded independent outcome assessor</li></ul>

DESIGN A		DESIGN B	
Study type	<u>Randomized controlled trial with randomization of manuscripts</u> Each manuscript is randomized to be peer reviewed by either : <ul style="list-style-type: none"><li>A peer reviewer asked to fill in the reporting guidelines checklist in addition to the usual process, if allocated to the intervention group.</li><li>A peer reviewer following the usual process, if allocated to the control group.</li></ul>	Study type	<u>Interrupted time series analysis</u> In interrupted time series studies, data are collected at multiple time points before and after an intervention in order to detect whether or not the intervention had a significantly greater effect than any underlying secular trend. <ul style="list-style-type: none"><li><b>Period 1: usual peer-review process.</b> In the first part of the study, peer reviewers follow the usual process of peer review.</li><li><b>Period 2: addition of an expert to the usual peer review process.</b> An assessment by an expert is systematically added to the usual peer review process.</li></ul>
Setting	A single biomedical journal	Setting	A single biomedical journal
Type of manuscript assessed by the peer reviewer	The manuscript(s) used in the study are the actual manuscripts submitted to the journal(s) and selected for peer-review during the time of the study (i.e. 2 years follow-up).	Type of manuscript assessed by the peer reviewer	The manuscript(s) used in the study are the actual manuscripts submitted to the journal(s) and selected for peer-review during the time of the study (i.e. 2 years follow-up).

- Series of vignette based studies
- 6 interventions
- Pairs of vignettes evaluated by 204 experts

Which design would you chose?

## What should be the outcome?

---

- Quality of the final manuscript (how is high quality defined, who is to decide?)
- Quality of the peer review report (editor's subjective assessment, validated scales to assess the quality of the peer review report)

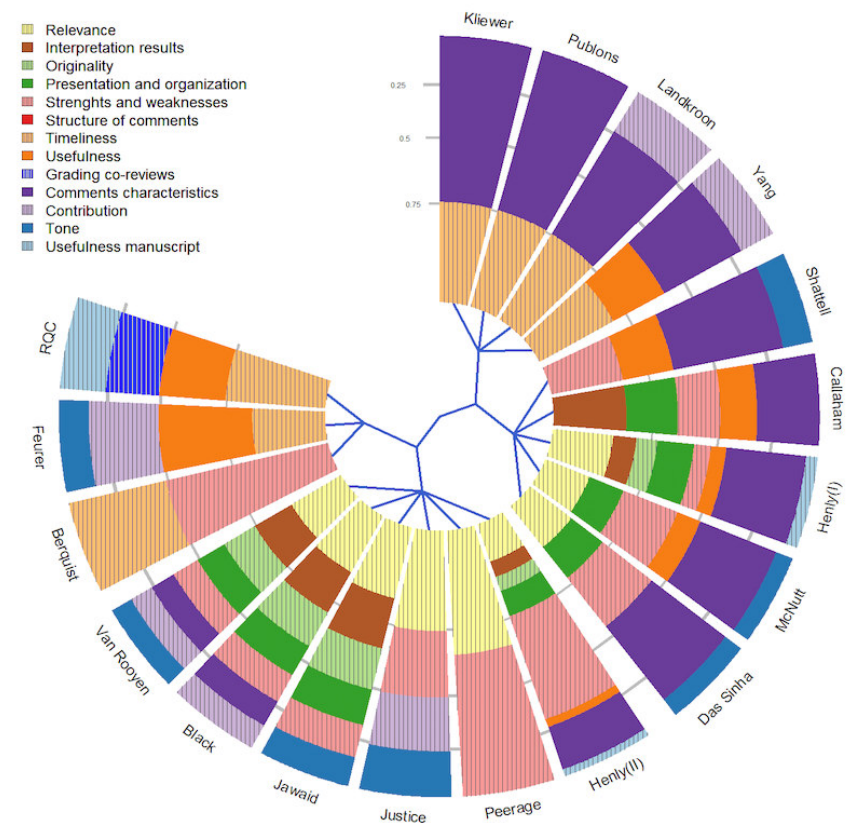


# What should be the outcome?

---

Systematic review of all tools assessing the quality of a peer-review report

- 23 scales
- 1 checklist
- No tool defined the concept of quality
- Number of domains varied



# Make peer review scientific

Thirty years on from the first congress on peer review, **Drummond Rennie** reflects on the improvements brought about by research into the process — and calls for more.

COMMENT

ILLUSTRATION BY DAVID PARKINS



# Acknowledgements

---

- Amytis Heim (master student)
- Anthony Chauvin (PhD student)
- Philippe Ravaud (Director of the Centre of Epidemiology Biostatistics Sorbonne Paris Cité)
- Cecilia Superchi (MiRoR PhD student)

---



# A 2-step peer-review system to improve completeness of reporting. Theoretical background of the intervention

