Improving peer review via controlled experiments

Isabelle Boutron

METHODS team
Research Centre of Epidemiology Biostatistics Sorbonne Paris Cité
Paris Descartes University

Background

Centre de Recherche Epidémiologie et Statistique Sorbonne Paris Cité
Inserm UMR 1153

METHODS Team

Methods of therapeutic evaluation of chronic diseases

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

UNIVERSITÉ PARIS DESCARTES

equator network

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

- 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

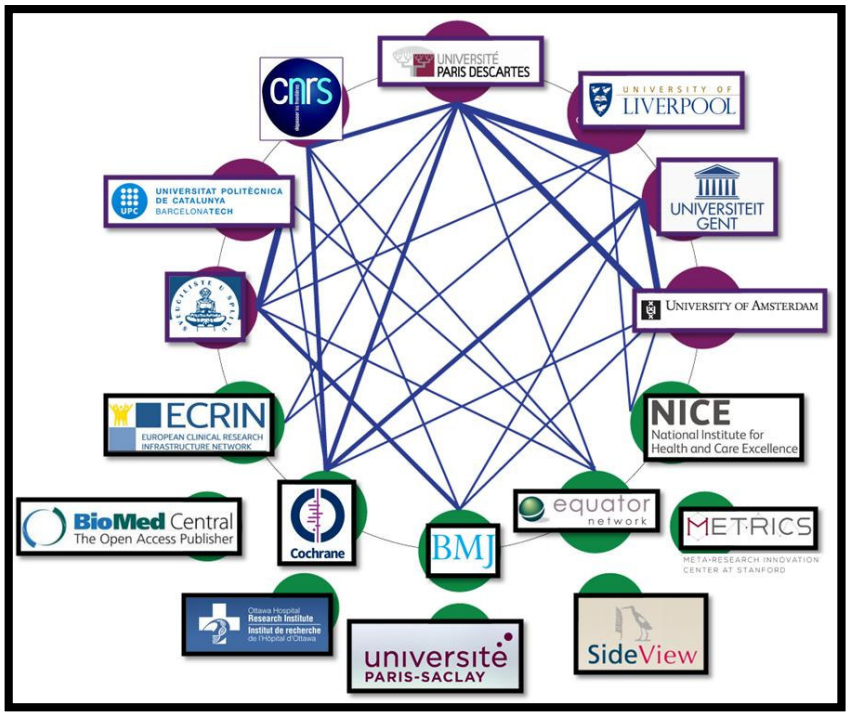
MiRoR project

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

Projects

-Planning
-Conduct
-Reporting
-**Peer review**

7 European Universities and 10 International Partners



MARIE CURIE ACTIONS

European Commission

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)

Research Letter

FREE

September 27, 2016

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

Kanu Okike, MD, MPH[1]; Kevin T. Hug, MD[2]; Mininder S. Kocher, MD, MPH[3]; 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

thebmj

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%).

Hopewell S, BMJ, 2014

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%

Hopewell S, BMJ, 2014

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

PLOS | MEDICINE

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

Riveros Plos Med, 2013

Detection of selective reporting of outcomes

OPEN ACCESS Freely available online

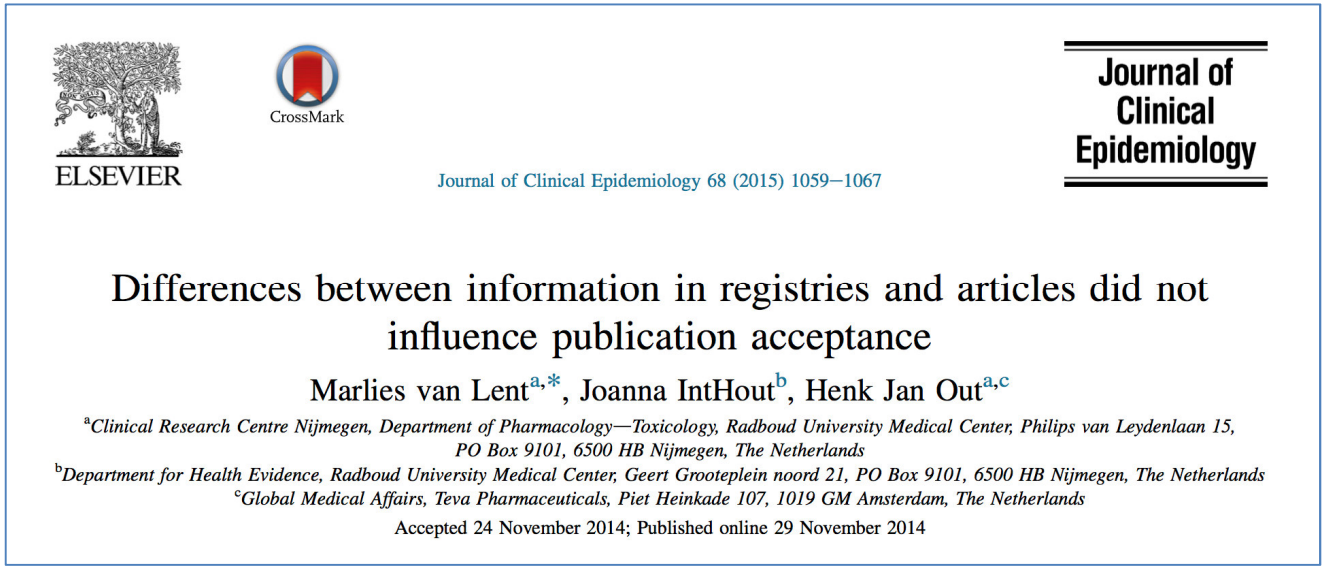
PLOS ONE

Use of Trial Register Information during the Peer Review Process

Sylvain Mathieu[1,2], An-Wen Chan[3], Philippe Ravaud[1]*

- 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



ELSEVIER

CrossMark

Journal of Clinical Epidemiology 68 (2015) 1059–1067

Journal of Clinical Epidemiology

Differences between information in registries and articles did not influence publication acceptance

Marlies van Lent[a,*], Joanna IntHout[b], Henk Jan Out[a,c]

[a]*Clinical Research Centre Nijmegen, Department of Pharmacology—Toxicology, Radboud University Medical Center, Philips van Leydenlaan 15, PO Box 9101, 6500 HB Nijmegen, The Netherlands*

[b]*Department for Health Evidence, Radboud University Medical Center, Geert Grooteplein noord 21, PO Box 9101, 6500 HB Nijmegen, The Netherlands*

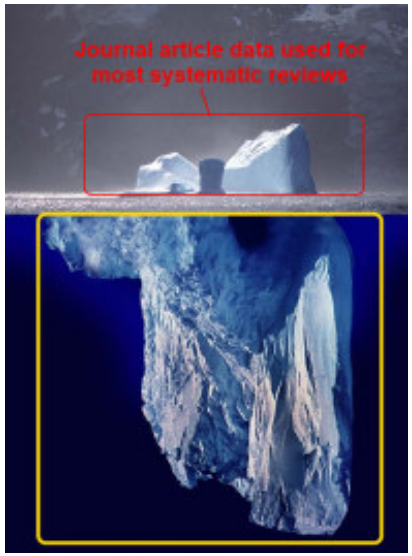
[c]*Global Medical Affairs, Teva Pharmaceuticals, Piet Heinkade 107, 1019 GM Amsterdam, The Netherlands*

Accepted 24 November 2014; Published online 29 November 2014

- 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

Boutron JAMA 2009

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).

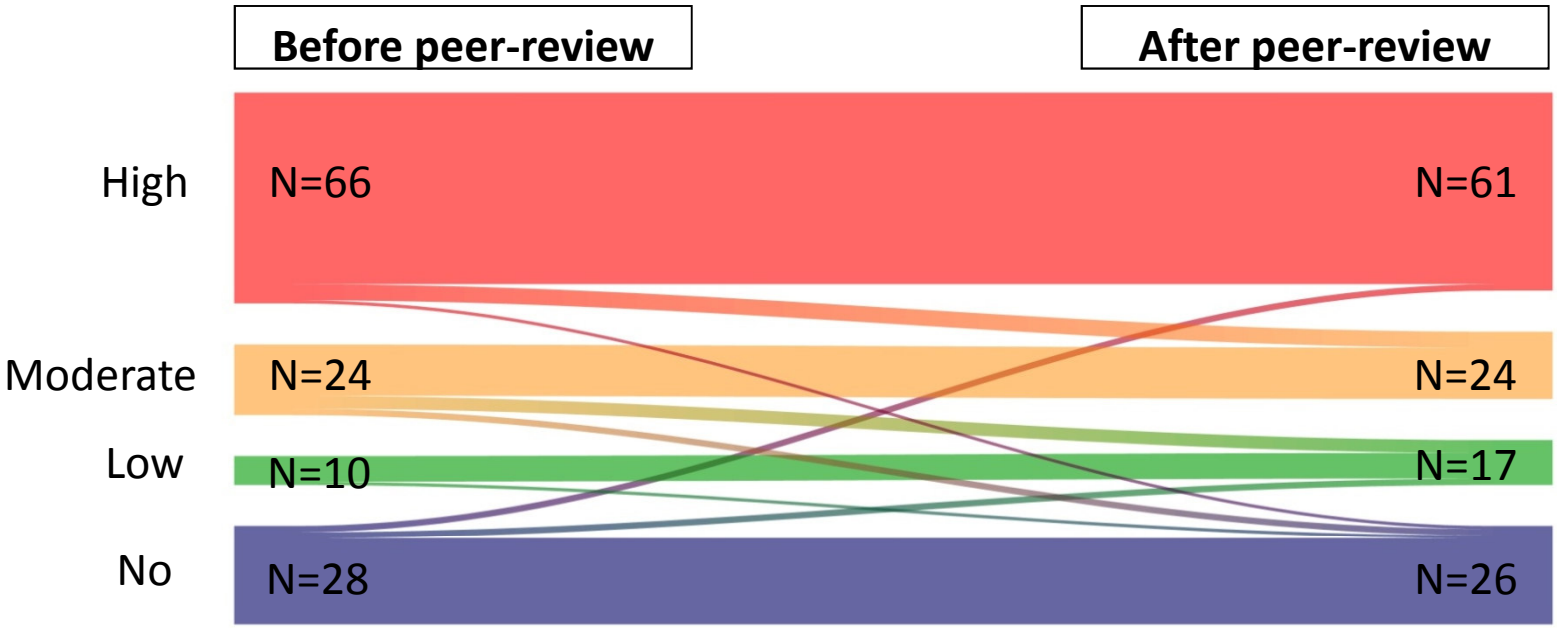
- 55% of submitted manuscripts, peer reviewers identified at least one example of spin

- 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

Lazarus, J Clin Epidemiol. 2016

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



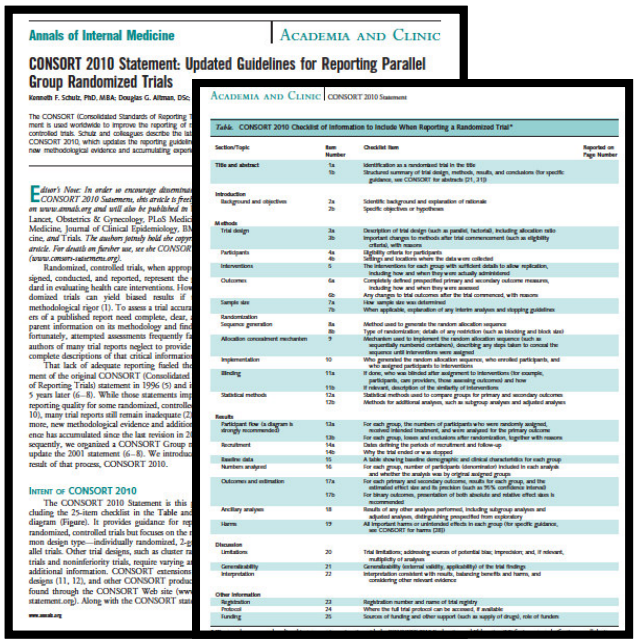
76% Peer reviewers failed to identify spin in abstract conclusions

Lazarus, J Clin Epidemiol. 2016

Controlled experiments to improve the peer-review process

Interventions to improve transparency

CONSORT Statement, recommandations for reporting RCTs


Annals of Internal Medicine | Academia and Clinic

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

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

Academia and Clinic | CONSORT 2010 Statement

Table. CONSORT 2010 Checklist of Information to Include When Reporting a Randomized Trial*

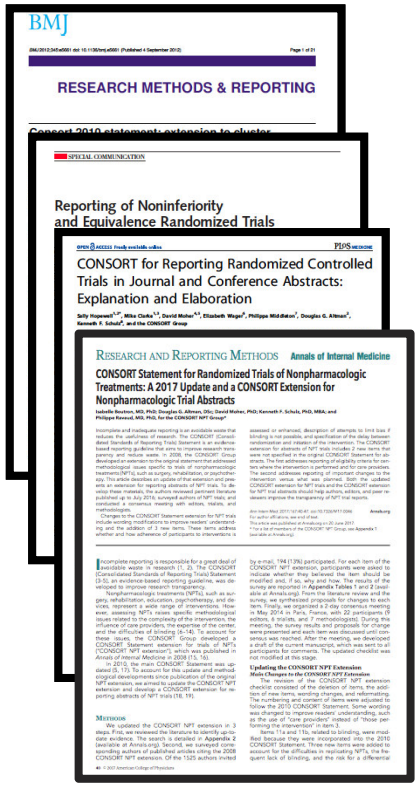

RESEARCH METHODS & REPORTING

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

David Moher,[1] Sally Hopewell,[2] Kenneth F Schulz,[3] Victor Montori,[4] Peter C Gøtzsche,[5] P J Devereaux,[6] Diana Elbourne,[7] Matthias Egger,[8] Douglas G Altman[2]

ABSTRACT

Extensions of CONSORT


BMJ

RESEARCH METHODS & REPORTING

SPECIAL COMMUNICATION

Reporting of Noninferiority and Equivalence Randomized Trials

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

Research and Reporting Methods | Annals of Internal Medicine

CONSORT Statement for Randomized Trials 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**

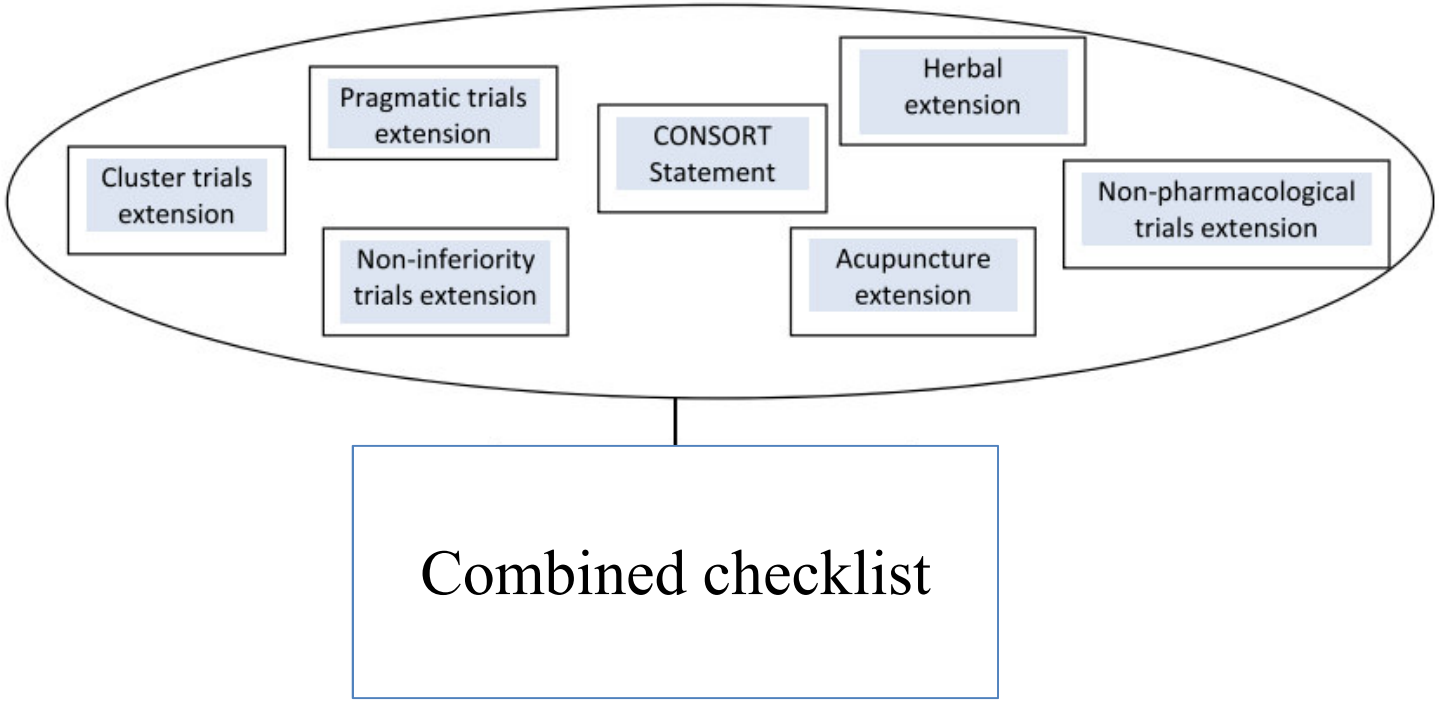
Hopewell *et al. BMC Medicine* (2016) 14:199
DOI 10.1186/s12916-016-0736-x

BMC Medicine

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



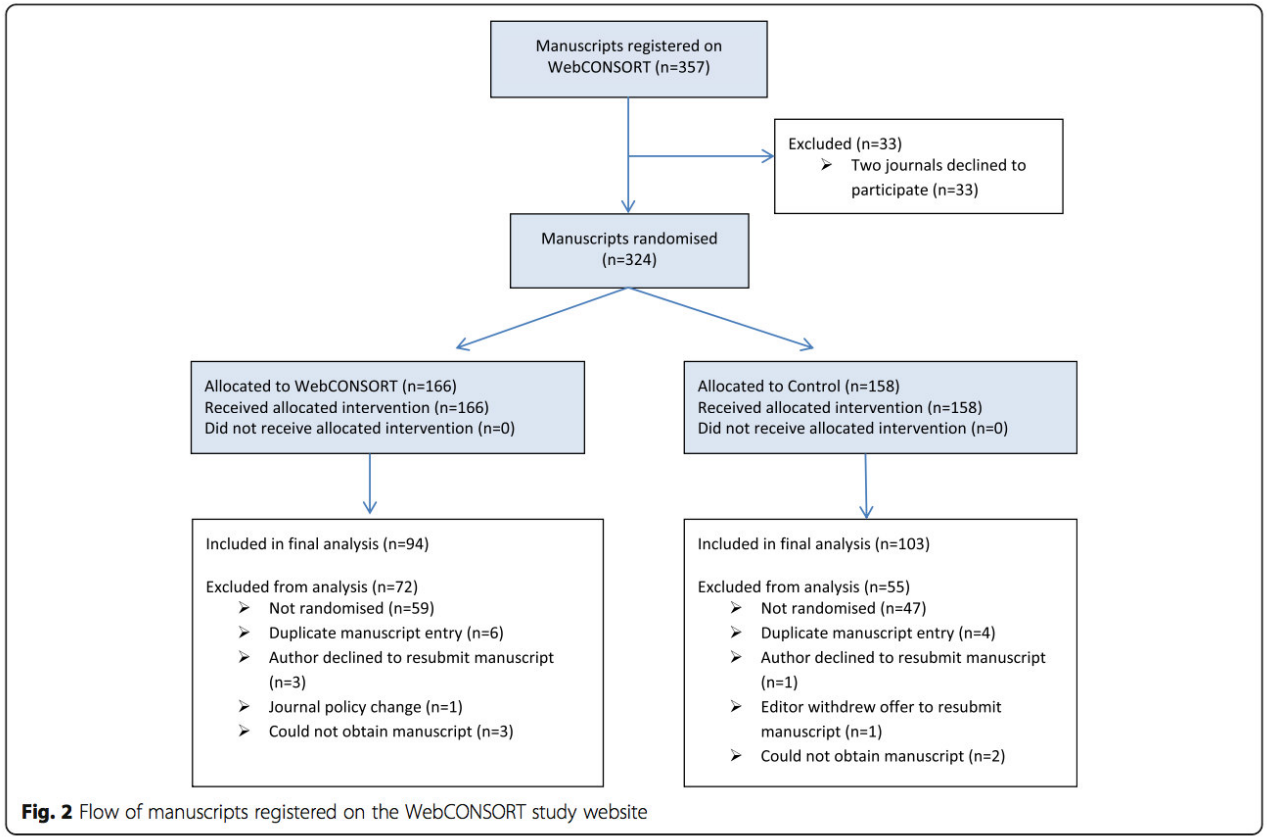


Fig. 2 Flow of manuscripts registered on the WebCONSORT study website

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

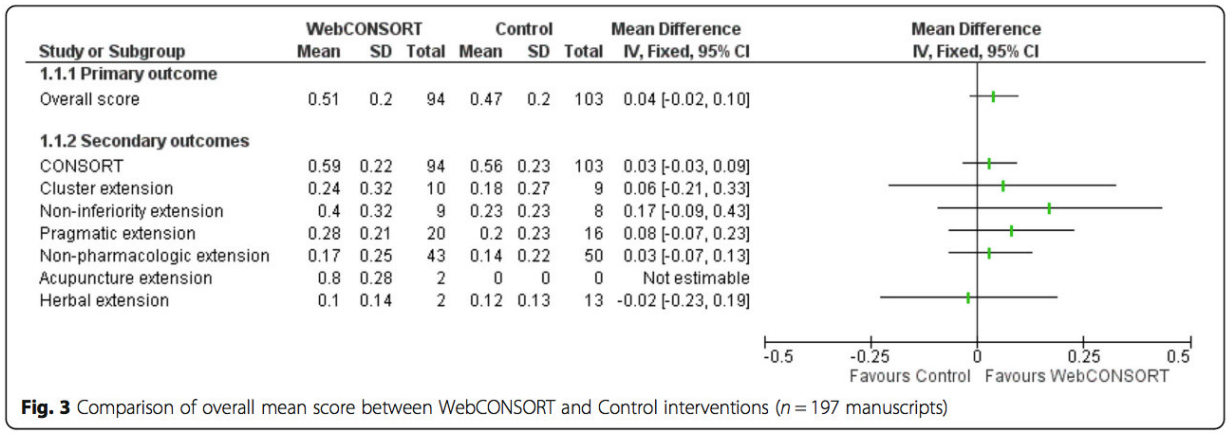


Fig. 3 Comparison of overall mean score between WebCONSORT and Control interventions ($n = 197$ manuscripts)

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

Blanco *et al. Trials* (2018) 19:80
DOI 10.1186/s13063-018-2475-0

Trials

COMMENTARY Open Access

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

David Blanco[1,3*], Alice M. Biggane[2,3], Erik Cobo[1] and MiRoR network

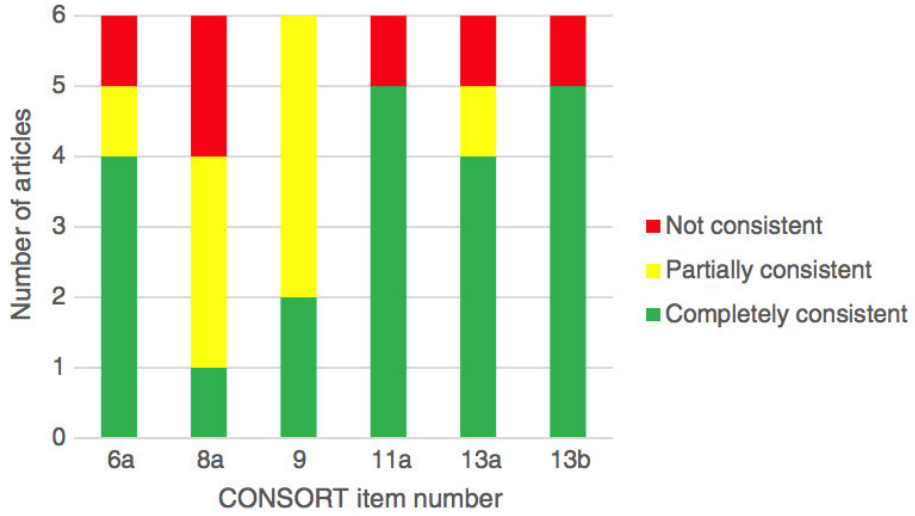


Fig. 1 Reporting inconsistencies found for the six papers that used the appropriate CONSORT checklist

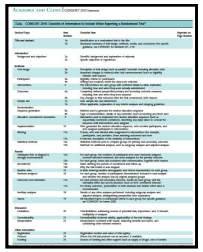
Format

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

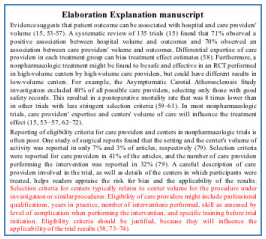
Format

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

Checklist



E & E



Item 6 a

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

Writing Aid Tool

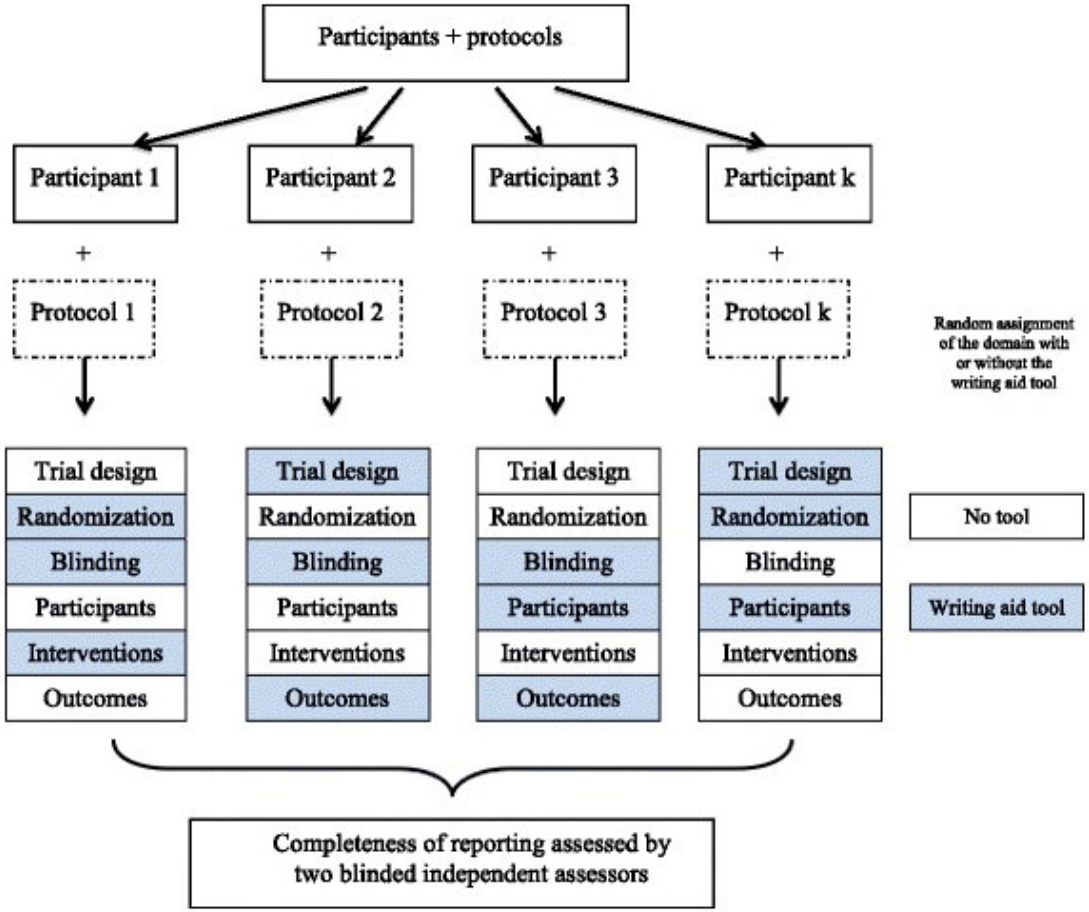
Expanded, combined and tailored checklist

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

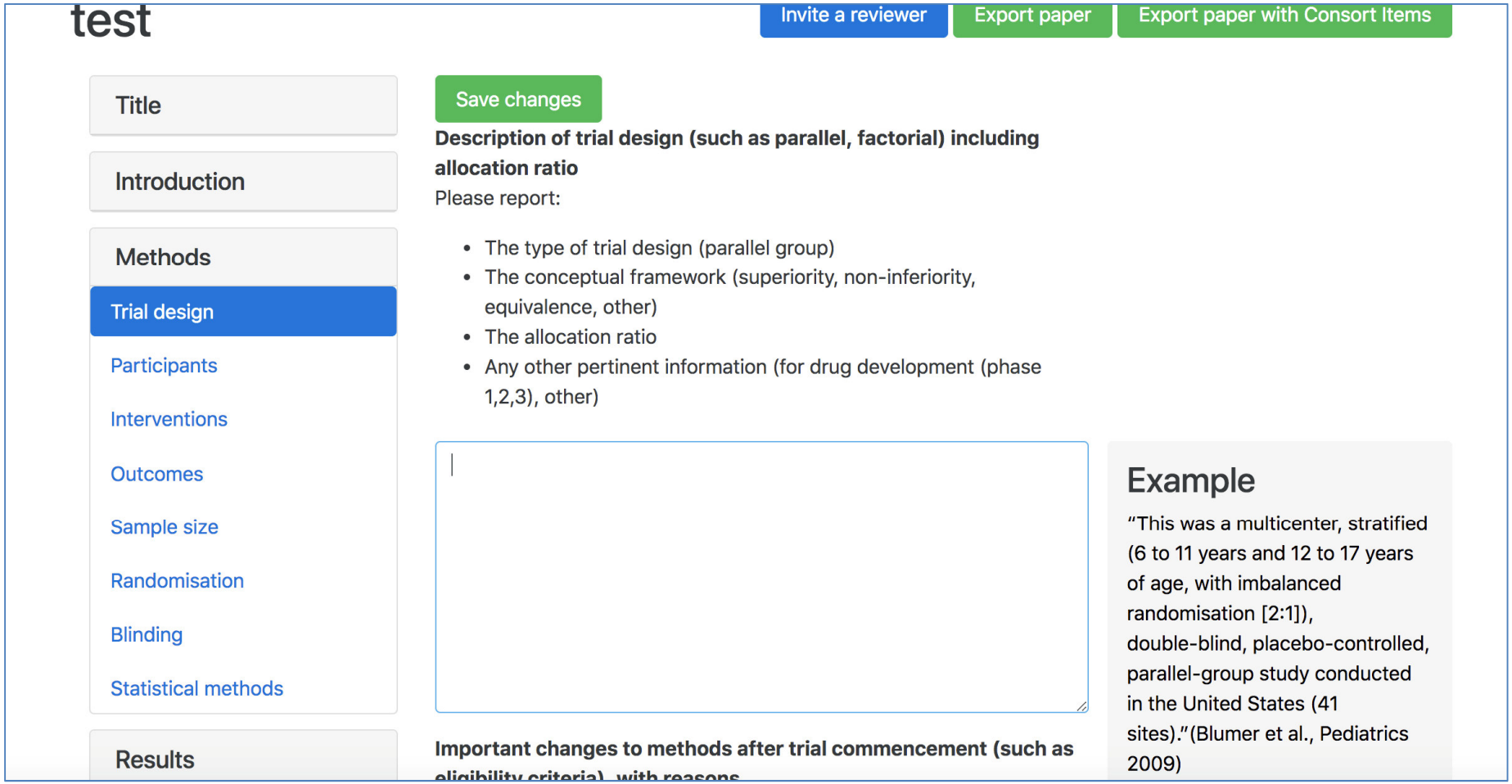
Mean Difference = 2.1 [1.5;2.7]; p <0.001

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

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

Barnes C, Boutron I, … Ravaud P. **BMC Med**. 2015

Cobweb


test
Invite a reviewer
Export paper
Export 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)
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)
Important changes to methods after trial commencement (such as

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

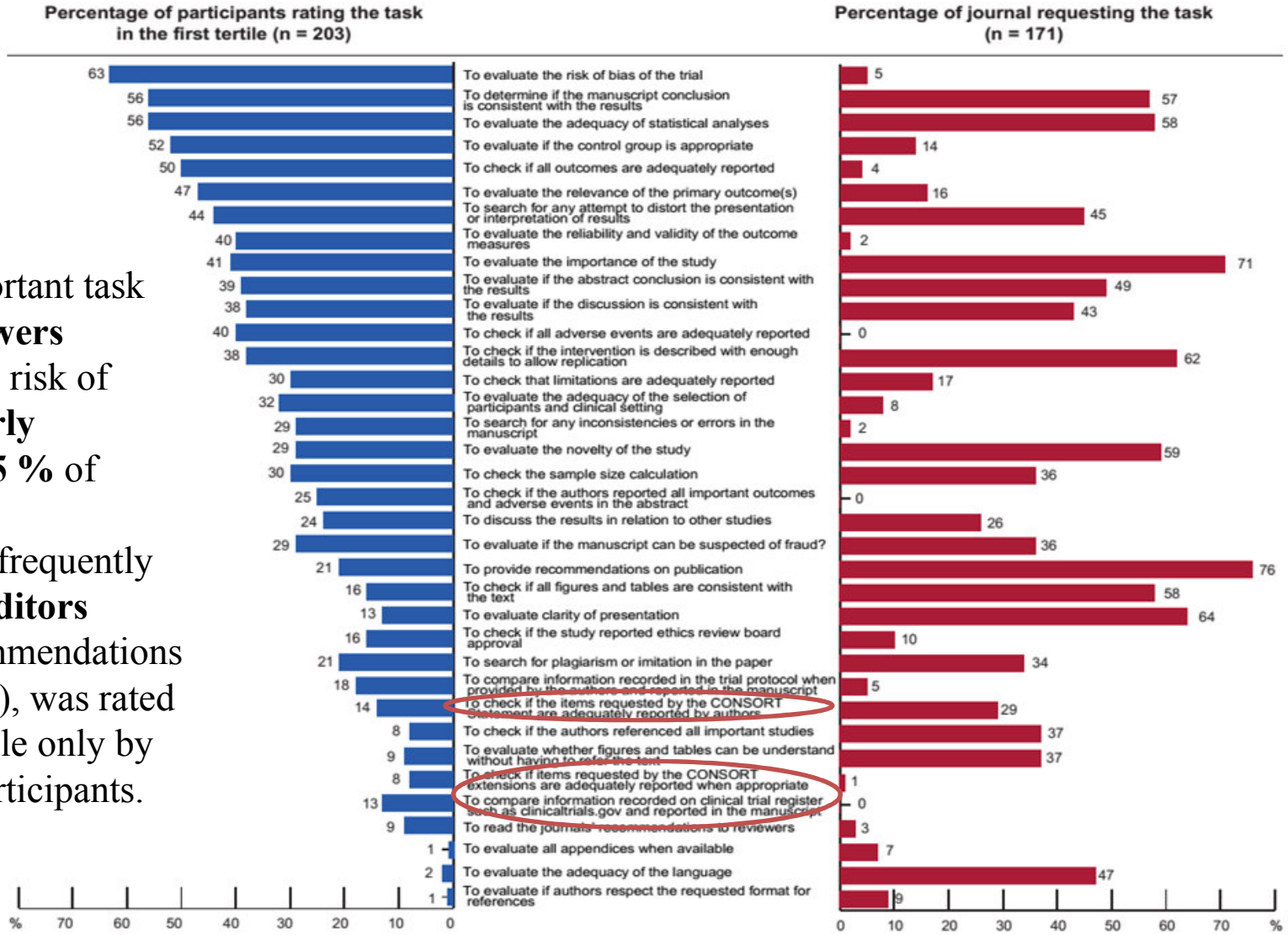
Anthony Chauvin[2,3†], Philippe Ravaud[1,2,3], Gabriel Baron[1,3], Caroline Barnes[2,3] and Isabelle Boutron[1,2,3*†]

- 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.



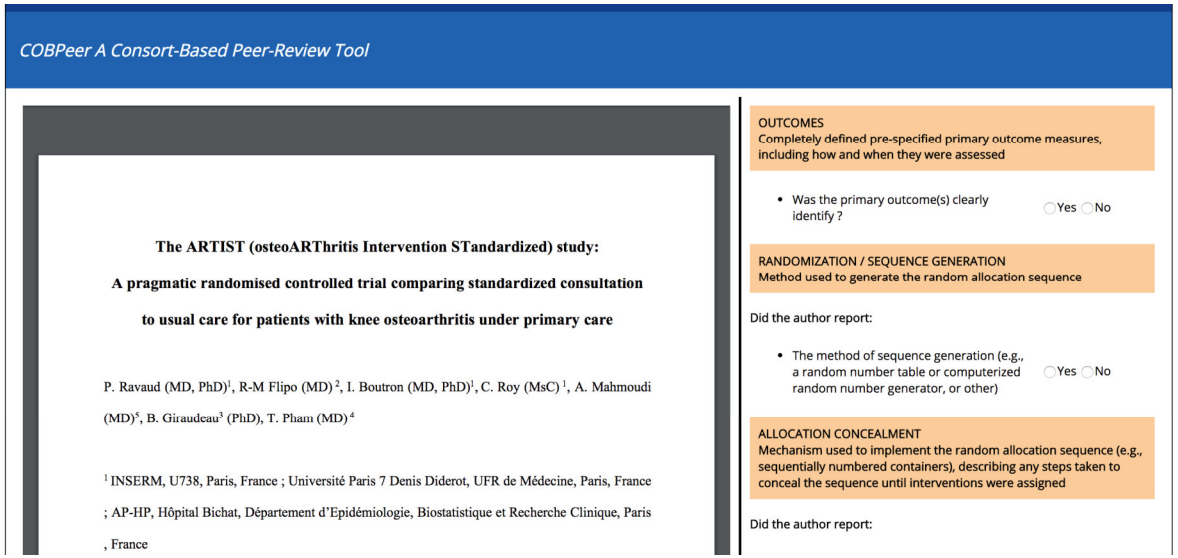
A 2-step peer-review system to improve completeness of reporting. Theoretical background of the intervention

- Peer-reviewers consider that assessing adherence to reporting guidelines and switch in outcomes is not their remit.

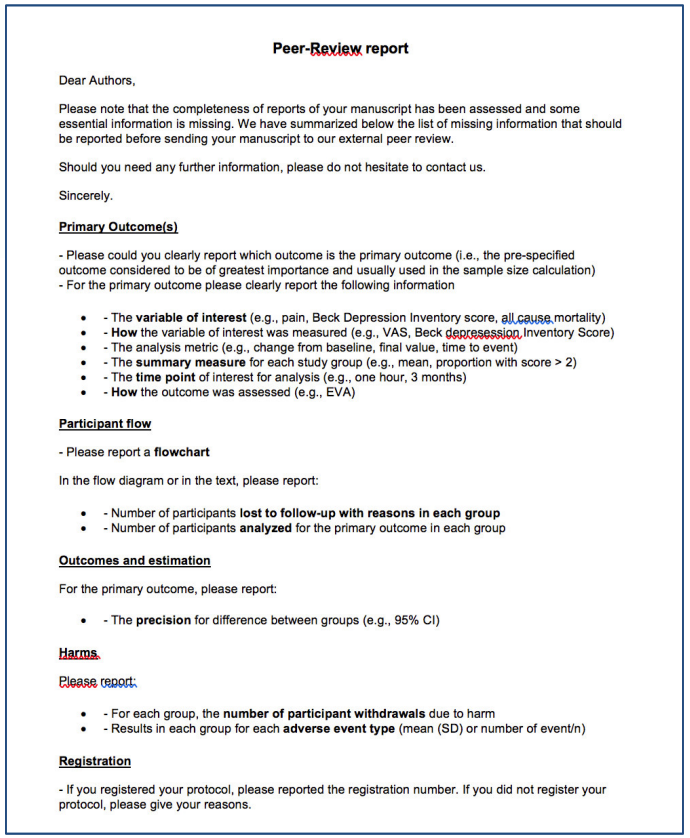
- Detection of inadequate reporting does not involve high level of expertise and could be performed by early career researchers who would gain expertise in peer-review.

- 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



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 depresession 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 ?	○	○

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

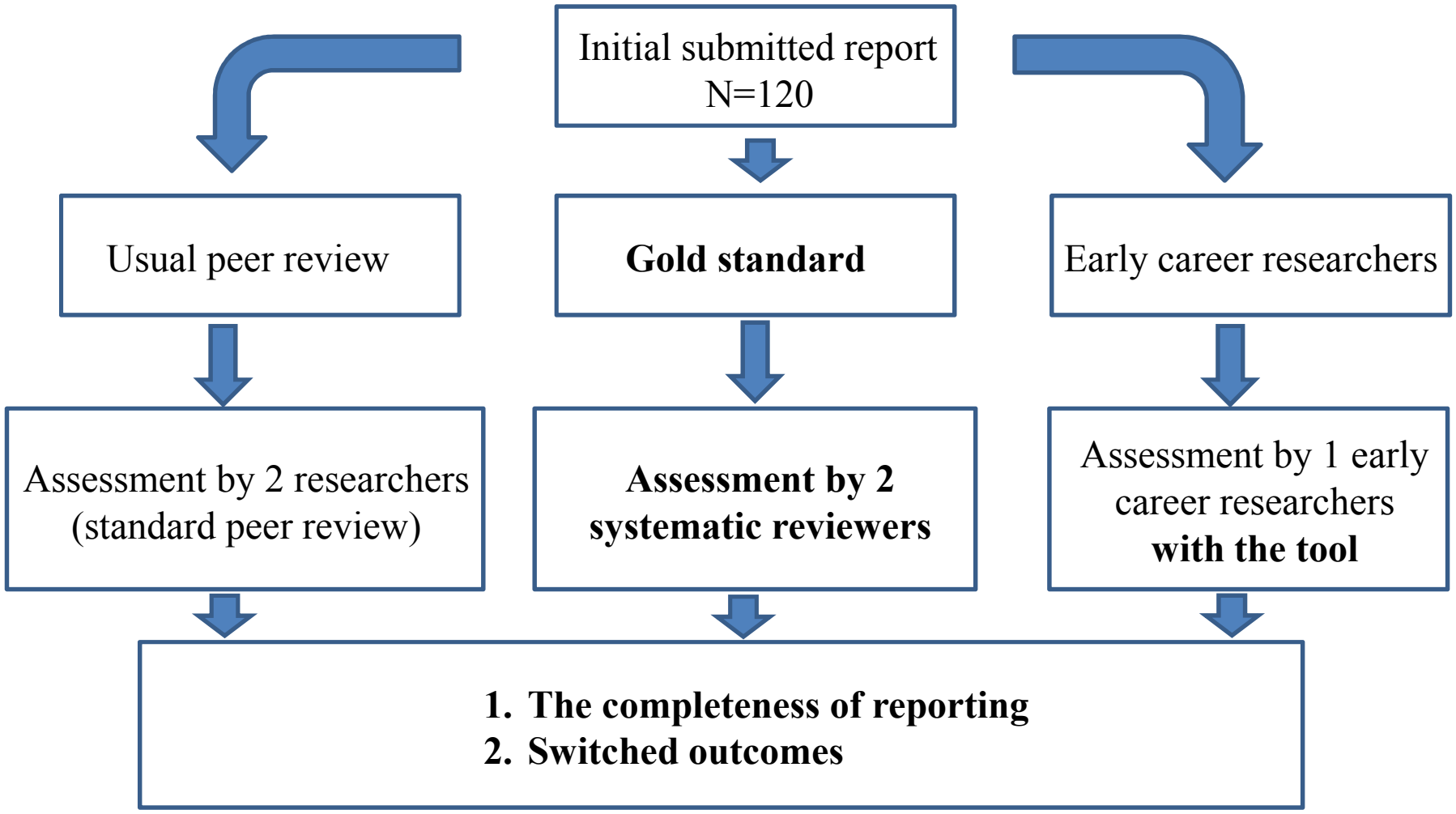
1/ The variable of interest: NOT REPORTED BY AUTHORS
2/ How the outcome(s) were assessed: NOT REPORTED BY AUTHORS
3/ The format of the primary outcome data: NOT REPORTED BY AUTHORS
4/ The summary measure for each study group: NOT REPORTED BY AUTHORS
5/ The time frame to measure primary outcome: NOT REPORTED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

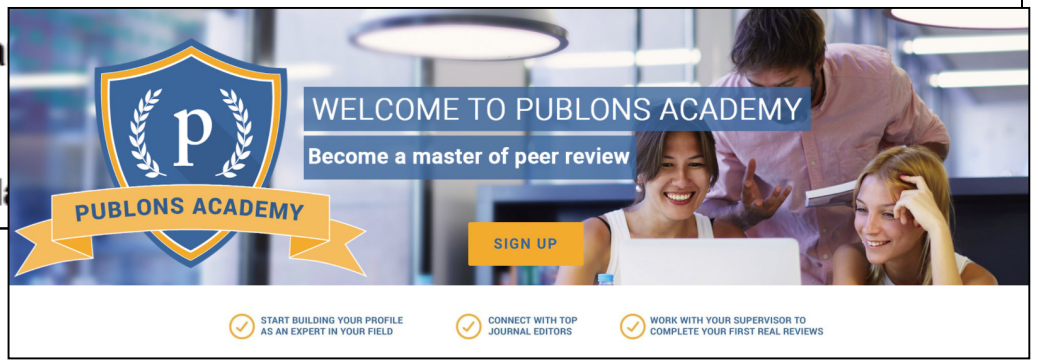
Button *et al. BMC Psychology* (2016) 4:59
DOI 10.1186/s40359-016-0167-7

BMC Psychology

EDITORIAL Open Access

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

Katherine S. Button[1*], Liz Bal[2], Anna Clark[2] and Tim Shipley[2]



nature COMMUNICATIONS

EDITORIAL OPEN

Towards an unbiased view of science

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

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*

statreviewer

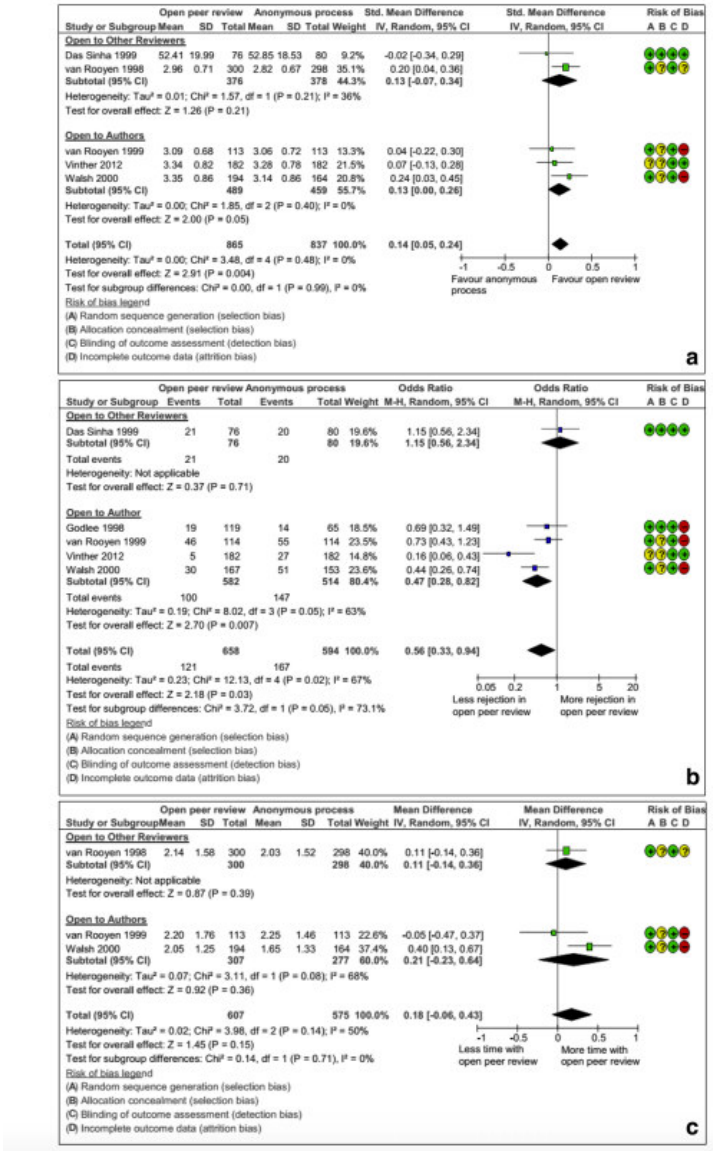
Could Robots Handle Peer Review?

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

What is the evidence?

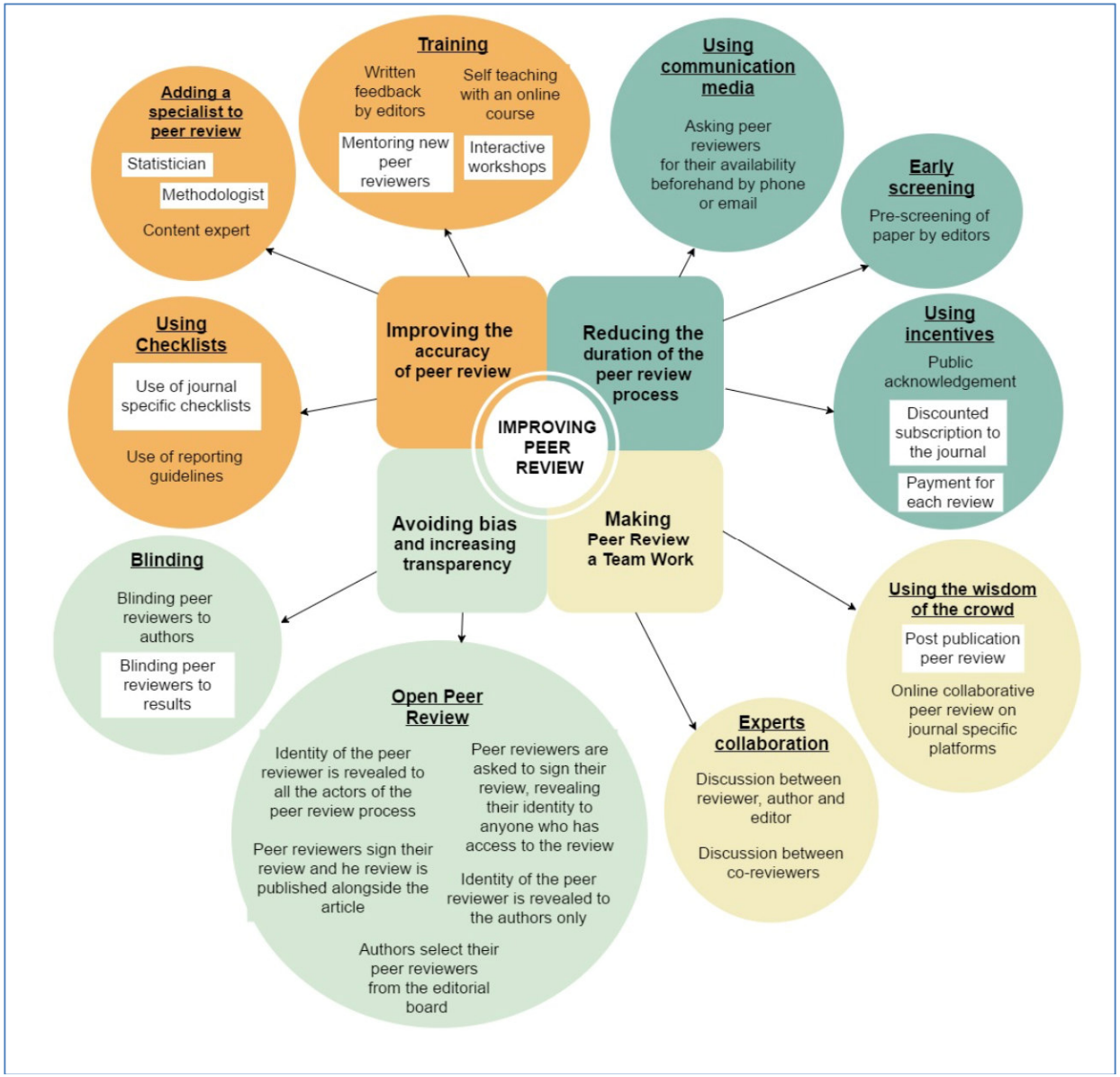
A Systematic Review and Meta-analysis of Interventions to Improve the Quality of Peer Review

- 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



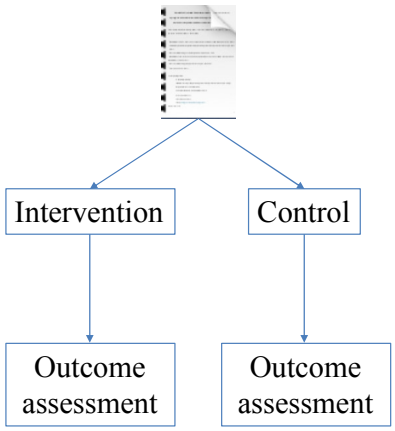
Bruce, BMC Med, 2016

What should be the study design?

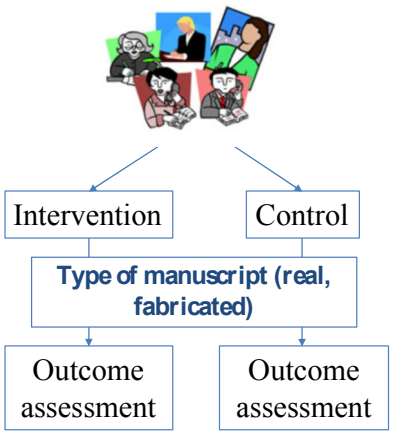


What should be the study design?

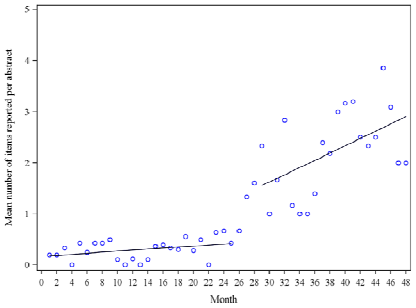
Manuscripts



Peer reviewers



Time series analysis

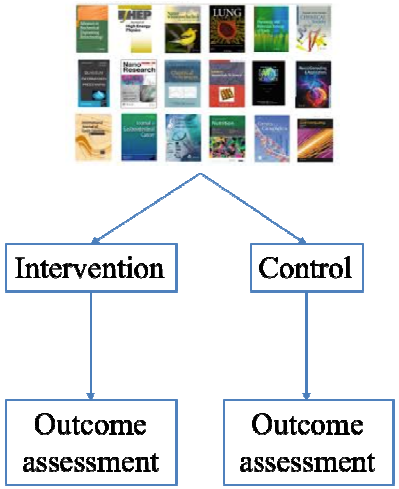


Active implementation
(Lancet, Ann Intern Med)

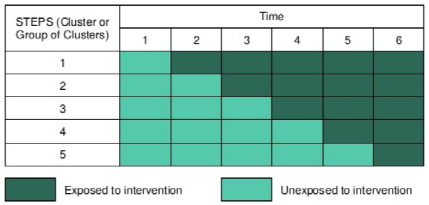
Level change: p = 0.0035
Trend change: p = 0.0154

Paired study design

Cluster RCTs



Stepped-wedge RCT



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

What should be the study design?

Fixed for each intervention

Does the use of reporting guidelines 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 • 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 • By a blinded independent outcome assessor

Varied within a same intervention

DESIGN A		DESIGN B	
Study type	**Randomized controlled trial with randomization of manuscripts** Each manuscript is randomized to be peer reviewed by either : • A peer reviewer asked to fill in the reporting guidelines checklist in addition to the usual process, if allocated to the intervention group. • A peer reviewer following the usual process, if allocated to the control group.	**Study type**	**Interrupted time series analysis** 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. • **Period 1: usual peer-review process.** In the first part of the study, peer reviewers follow the usual process of peer review. • **Period 2: addition of an expert to the usual peer review process.** An assessment by an expert is systematically added to the usual peer review process.
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)

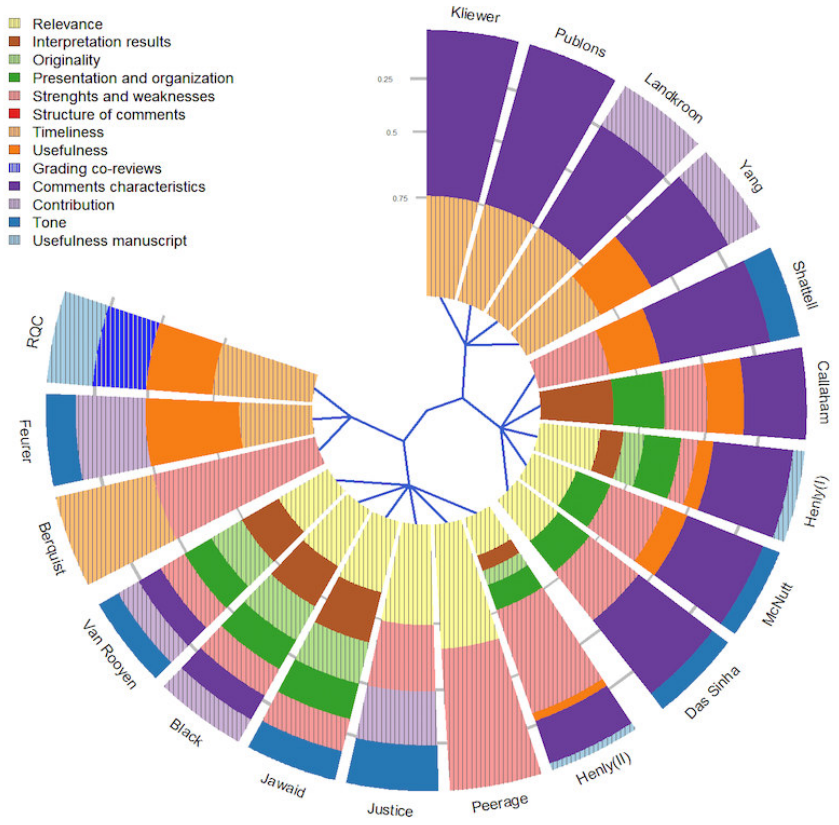
What should be the outcome?

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



Superchi C, PEERE, 2018

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

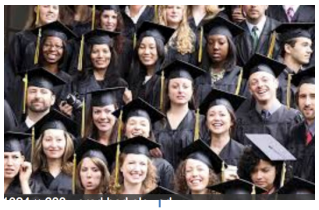
Acknowledgements

- 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

Authors



Impossible to adequately assess /lack of transparency

Assessment of a transparent manuscript

Assessment of a transparent manuscript