

Name that bias: lessons learned from empirical studies of bias in clinical research

Lisa Bero, Professor, Department of Clinical Pharmacy and Institute for Health Policy Studies, UCSF and Director SF Cochrane Center

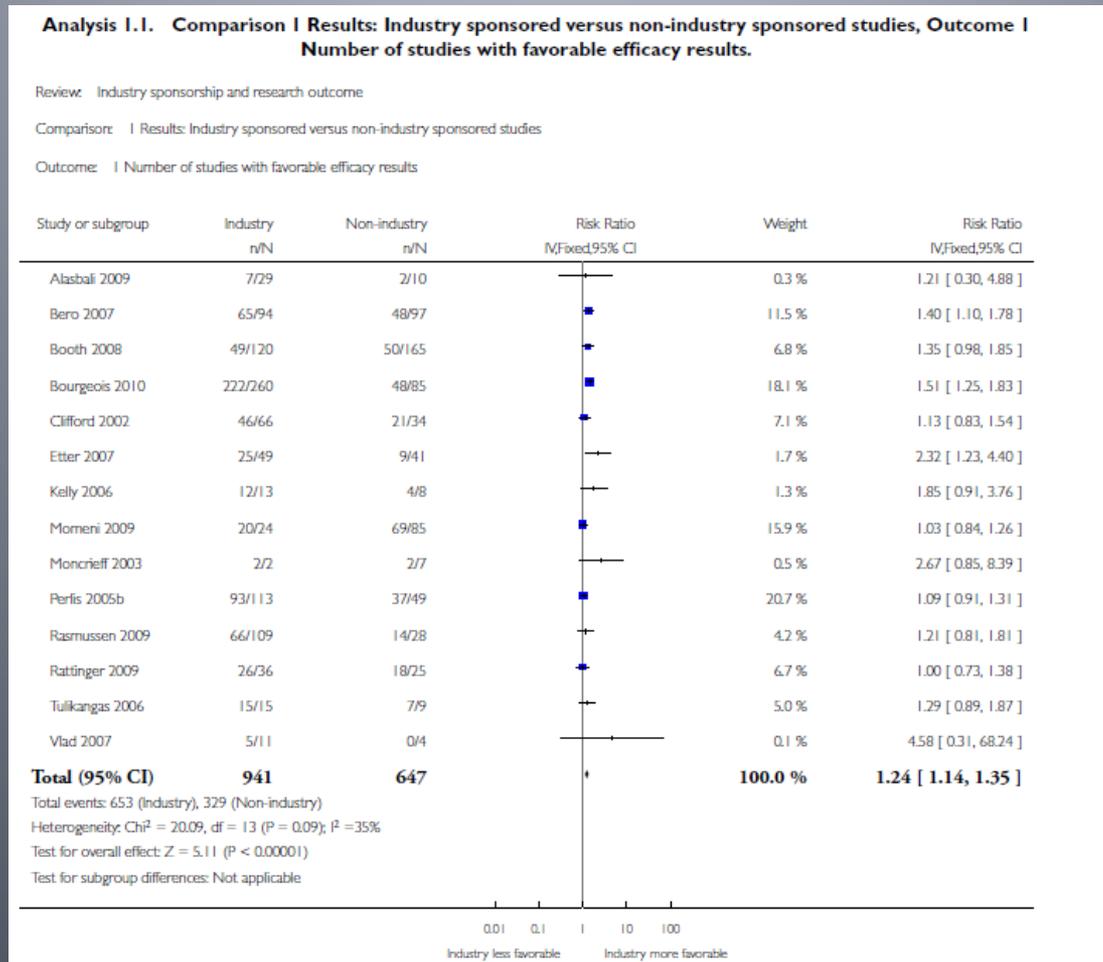
Risk of Bias

- Methodological characteristics of a study that can introduce a systematic error in the magnitude or direction of the results (Higgins and Green 2008).

Underlying principles

- The objective of pharmacological, toxicological, chemical studies is to determine if an intervention or an exposure is causally related to an outcome.
- Risk of bias criteria are applicable to all these types of studies as they are adherent to the same scientific principles.
- As the starting assumption is that there is no effect on outcome, the methodological criteria are independent of the “expected” or “desired” direction of the effect. In the case of drug efficacy, this may be a “positive” outcome (to observe an effect or larger effect size), in the case of drug harm this may be a “negative” outcome (to observe no effect or smaller effect size).
- The methodological issues are independent of the data stream (animal or human).

Industry sponsored drug studies are more likely to have favorable efficacy results than non-industry sponsored studies



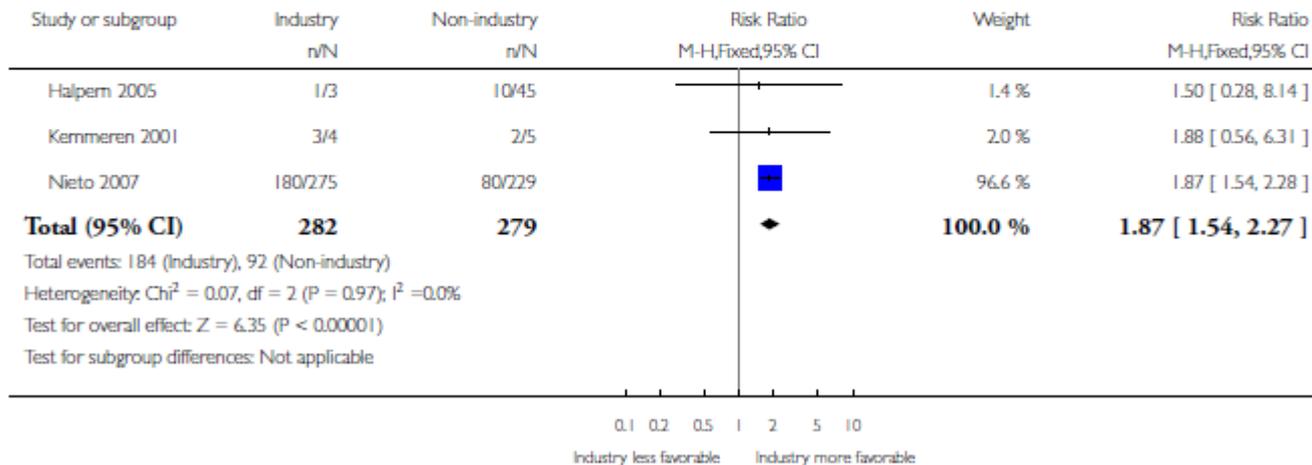
Industry sponsored drug studies are more likely to have favorable harm results than non-industry sponsored studies

Analysis 1.2. Comparison 1 Results: Industry sponsored versus non-industry sponsored studies, Outcome 2 Number of studies with favorable harms results.

Review: Industry sponsorship and research outcome

Comparison: 1 Results: Industry sponsored versus non-industry sponsored studies

Outcome: 2 Number of studies with favorable harms results



Risk of Bias

- In human clinical trials of drug efficacy, studies with a high risk of bias, such as those lacking randomization, allocation concealment, or blinding of participants, personnel and outcome assessors produce larger treatment effect sizes, thus falsely inflating the efficacy of the drugs, compared to studies that have these design features (Schulz et al. 1995; Schulz and Grimes 2002a, b).
- Biased human studies assessing the harms of drugs are less likely to report statistically significant adverse effects (Nieto et al. 2007).

Risk of Bias is not.....

- An assessment of how the study is conducted (eg, in compliance with human subjects guidelines)
- As assessment of how the study is reported (eg, study population described, abstract is structured).
- The same as imprecision. While bias refers to systematic error, imprecision refers to random error. Although smaller studies are less precise, they may not be more biased.

Types of bias

Selection bias

- Introduces systematic differences between baseline characteristics in comparison groups
- Minimized by randomization and concealment of allocation

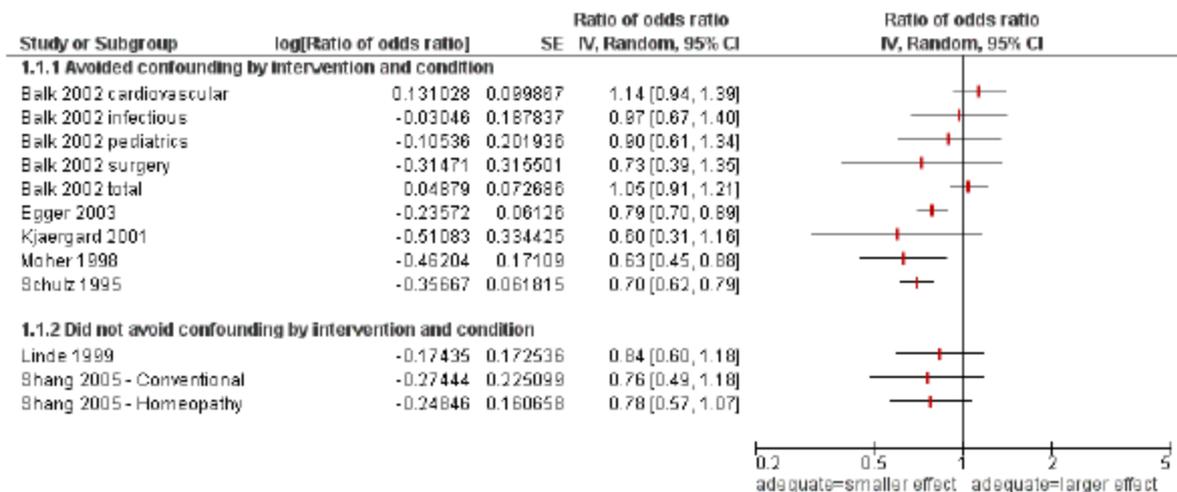
Empirical evidence of bias

- Analysis of 250 randomized controlled trials
- Inadequate concealment of allocation
 - Estimates of effect 30-41% greater than studies with adequate concealment
- Inadequate generation of allocation schedule
 - Small change in estimate of treatment effect

- Schulz KF, Chalmers I, Hayes RJ, Altman DG. 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273(5): 408-412.

Effect sizes of studies with adequate concealment of allocation differ from those with inadequate concealment of allocation

Figure 2. Studies of controlled trials with adequate concealment of allocation compared with inadequate/unclear concealment of allocation across different interventions and conditions - ratio of odds ratios



Citation: Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews* 2011, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.pub3.

Performance / detection biases

- *Performance* - systematic difference between treatment and control groups with regard to care or other exposure besides the intervention / treatment.
- Blinding of investigators can protect against performance bias
- *Detection* - systematic differences between treatment and control groups with regards to how outcomes are assessed.
- Blinding of outcome assessors is a primary way of reducing detection bias.

Empirical evidence of bias

- No double blinding
 - Analysis of 250 randomized controlled trials
 - Estimates of effect 17% greater than studies with double blinding
 - Schulz KF, Chalmers I, Hayes RJ, Altman DG. 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273(5): 408-412.

Who is blinded and how?

- “double blinding” not informative.
- Need to know:
 - Participants
 - Investigators
 - Data collectors
 - Outcome assessors
 - Data analysts
- Ways to achieve blinding in animal studies – having coded data analyzed by a statistician who is independent of the study team

Attrition / Exclusion bias

- The systematic difference between treatment and control groups in the number of subjects that were included in and completed the study.
- Data on whether all subjects in the study are accounted for and use of intention-to-treat analysis can reduce exclusion bias.

Attrition / exclusion bias

- Participants lost to follow-up may differ from those who complete the study
- Data could be differentially excluded from comparison groups
- Intention to treat analysis
 - Means analyzing all patients in the group to which they were assigned, regardless of whether they actually received the treatment / exposure and how else they may differ

Methods specific to the research question

- For drug studies, choice of dose / duration of treatment
- Also applies to toxicology studies
- “Intensity of intervention” applies to all studies

Gaming the dose

- High dose of drug A compared to low dose of drug B to “demonstrate” efficacy of A
- Low dose of drug A compared to high dose of drug B to “demonstrate” drug A has less adverse effects
- = favorable outcome for drug A

Gaming the dose

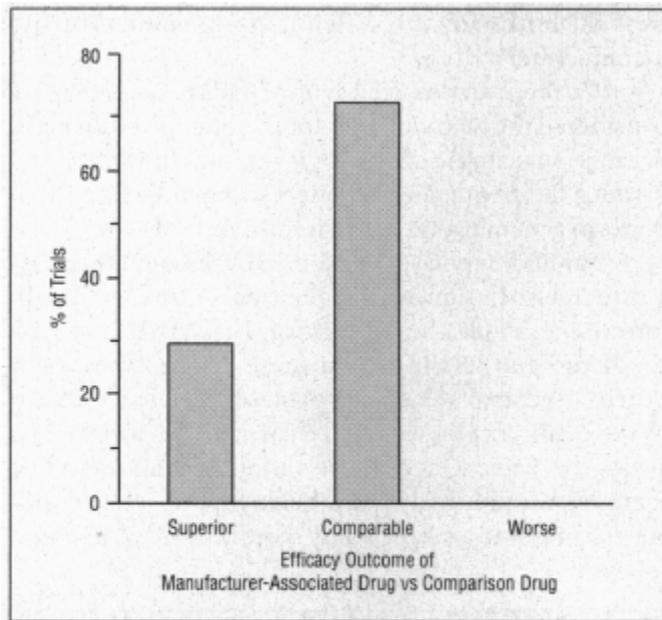


Figure 2. Efficacy performance of the manufacturer-associated drug relative to the comparison drug, based on the narrative interpretation of the study results. The percentage of all trials in which the manufacturer-associated drug was declared superior to, comparable with, or worse than the comparison drug in terms of efficacy outcome is shown.

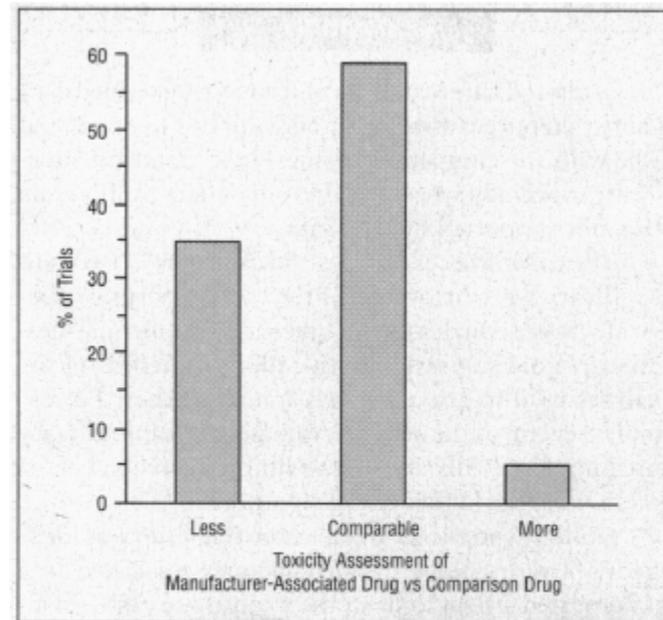
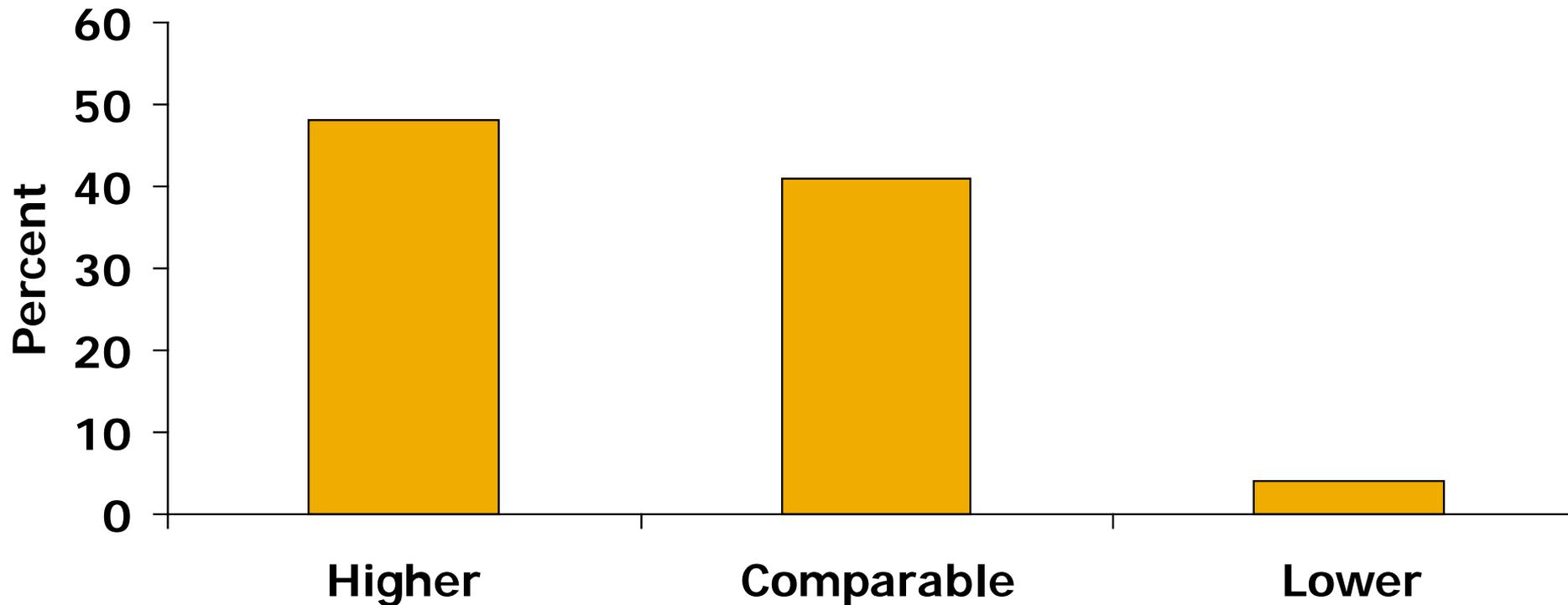


Figure 3. Toxicity assessment of the manufacturer-associated drug relative to the comparison drug, based on the narrative interpretation of the study results. The percentage of all trials (n=54) in which the manufacturer-associated drug was declared either less toxic, of comparable toxicity, or more toxic is shown. Two trials were excluded because they provided no toxicity information in the narrative interpretation of the study results.

Rochon et al., A study of manufacturer-supported trials of nonsteroidal Anti-inflammatory Drugs in the Treatment of Arthritis. Arch Intern Med. 1994; 154: 157-163

How did they get this result for efficacy?



**Dosage of Manufacturer's Drug vs.
Comparator Drug**

Rochon, Arch Intern Med 1994; 154: 157-63.

Selective reporting bias

- The selective reporting of entire studies (publication bias) or outcomes from studies (selective outcome reporting).
- Publication bias minimized by registries, published protocols.
- Selective reporting minimized by full access to study protocols, results, reports

Selective reporting of drug studies

Identify and characterize discrepancies, if any, between clinical trial data submitted to the Food and Drug Administration (FDA) in approved new drug applications (NDAs) and the corresponding published trials

- Rising, K, Bacchetti, P, and Bero, L. Reporting bias in drug trials submitted to the Food and Drug Administration . *PLoS Medicine*, 2008; 5 (11) e217 doi:10.1371/journal.pmed.0050217.
- Hart, B, Lundh, A and Bero, L. The effect of reporting bias on meta-analyses of drug trials: Re-analysis of meta-analyses. *BMJ*, 2011;343:d7202. doi: 10.1136/bmj.d7202

Some trials are not published

Of 164 Trials submitted in NDAs...

PUBLISHED within 5 years: 78% (128)

OF 33 NDAs...

All trials published: 52% (17)

NO trials published: 2 (with a total of 5 trials)

Predictors of Publication

Trial Characteristic	OR (95% CI)	p value
Favorable primary outcome (s)	4.77 (1.33-17.06)	0.018
Active control (vs. placebo only)	3.37 (1.02-11.22)	0.047

Papers include more outcomes favoring the test drug

- 179 primary outcomes reported in NDAs
 - 41 were omitted from the papers
- Papers had 138 outcomes also reported in NDAs (77%)
 - PLUS 15 additional outcomes that favored the test drug
 - PLUS 2 other neutral outcomes

Statistical Significance of Reported Outcomes Changed

- 43 outcomes in the NDAs did not favor the test drug
 - 20 were not included in the papers
 - 5 changed statistical significance, with 4 changing to favor test drug in the paper
- * changes in outcomes occurred in 36 (22%) trials found in 19 (58%) NDAs

Quality “scores”

How do we measure “quality”?

25 scales, 9 checklists and more.....

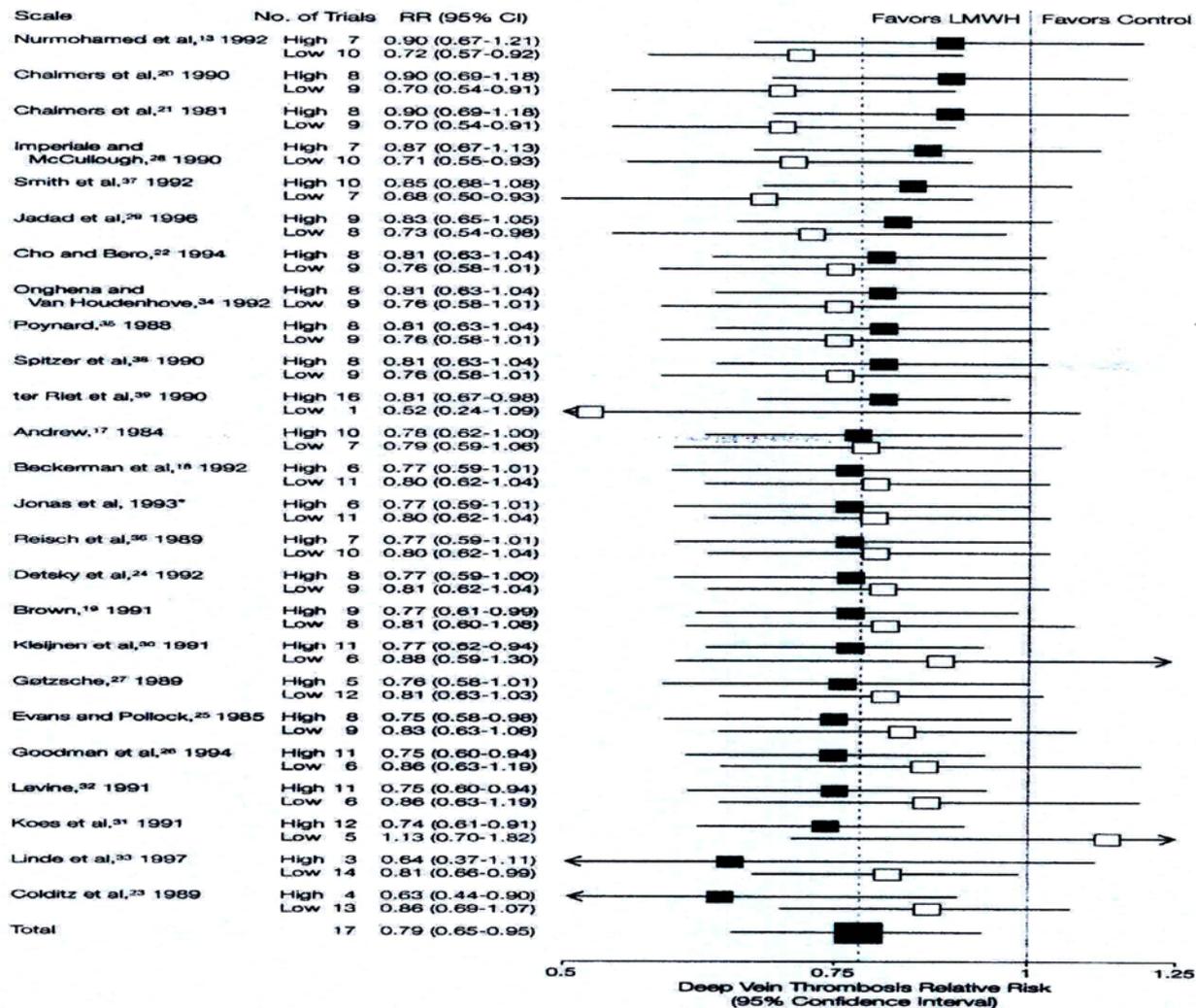
Table 1 Descriptive characteristics of published and unpublished scales used to assess the quality of randomized controlled trials (RCTs)

Scale Name ^a	Type of scale ^b	Quality defined ^c	Type of quality assessed ^d	Items selected ^e	Patient assignment ^f	Masking ^g	Patient follow-up ^h	Statistical analysis ⁱ	Number of items	Scale development ^j	Inter-rater reliability ^k	Time to complete ^l	Scoring range ^m	Detailed Instructions for scoring items ⁿ	Meta-analysis scores ^o
Andrew ^{3,44}	s	n	r	ac	y	y	n	y	11	nr	0.95 ^a	10	0-22	y	56
Annals ⁴	g	y	r	ac	y	y	n	y	34	nr	0.12 ^a	15	34-170	y	75
Beckerman ⁵	s	y	m	ac	y	y	y	y	25	nr	nr	10	0-25	n	35
Brown ^{6,45}	s	n	m	ac	y	n	n	n	6	nr	0.89 ^a	10	0-21	y	55
Chalmers, I ^{7,46}	g	y	m&r	ac	y	y	n	y	3	nr	nr	<10	0-9	y	58
Chalmers, TC ^{2,47}	g	n	m	ac	y	y	y	y	27	nr	nr	45	0-100	y	45
Cho ⁸	g	y	m&r	ac	y	y	n	y	24	nr	0.89 ^a	30	0-1	y	60
Colditz ⁹	g	n	m&r	ac	y	n	y	y	8	nr	nr	10	0-8	n	56
Detsky ¹⁰	g	n	m&r	ac	y	y	n	y	5	nr	nr	10	0-15	n	nr
Evans ¹¹	g	n	m&r	ac	y	y	n	y	33	nr	nr	15	0-100	n	63
Gotzsche ¹²	s	n	m&r	ac	y	y	n	y	8.8	nr	nr	15	0-8	n	25,38
Imperiale ¹³	g	n	m	ac	y	n	n	n	5	nr	0.79 ^a	<10	0-5	n	82
Jadad ¹⁴	g	y	r	pool	y	y	n	y	6	y	0.65,0.75 ^a	<10	0-8	y	56
	g	y	r	pool	y	y	n	n	3	y	0.66,0.77 ^a	<10	0-5	y	54
Jonas ⁹	g	n	m	ac	y	y	y	y	20	nr	0.6 ^a	20	0-100	y	nr
Kleijnen ¹⁵	g	p	m&r	ac	y	y	n	y	7	nr	0.87 ^a	15	0-100	y	42
Koes ¹⁶	s	p	m	ac	y	y	y	y	17	nr	0.8 ^a	15	0-100	y	37
Linde ⁴	g	y	m&r	ac	y	y	y	y	24	nr	nr	30	0-100	y	nr
Nurmohamed ¹⁷	s	p	m&r	ac	y	y	n	n	8	nr	nr	10	0-8	y	nr
Onghenia ¹⁸	s	n	m&r	ac	n	n	n	y	10	nr	nr	15	0-10	y	55
Poynard ^{19,48}	g	p	m&r	ac	n	y	y	y	14	nr	>.66 ^a	10	-2-26	y	13
Reisch ^{20,49}	g	n	m&r	ac	y	y	y	y	34	nr	0.99,0.71 ^a	30	0-34	y	45
Smith ²¹	s	n	m	ac	y	n	y	n	8	nr	nr	10	0-40	n	62
Spitzer ²²	s	n	m	ac	y	y	n	y	5	nr	nr	25	0-5	n	nr
Taps ²³	g	n	m&r	ac	y	y	y	y	29	nr	nr	30	0-100	y	nr
Ter Riet ²⁴	s	p	m&r	ac	y	y	y	y	18	nr	0.93 ^a	15	1-100	y	47

Several problems with all measures

- Mix reporting and actual study design
- No empirical evidence for weighting the scores
- Reliability and validity not measured
- Does not typically assess: the question asked, how the study is conducted, whether the study is reported (publication bias, selective outcome reporting)
- What does a quality “score” mean? How can you use them?

Figure 1. Results From Sensitivity Analyses Dividing Trials in High- and Low-Quality Strata, Using 25 Different Quality Assessment Scales



Relative risks (RRs) for deep vein thrombosis with 95% confidence intervals (CIs) are shown. LMWH indicates low-molecular-weight heparin. Black squares indicate estimates from high-quality trials and open squares indicate estimates from low-quality trials. Arrows indicate that the values are outside the range of the x axis. Broken line indicates combined estimate from all 17 trials. Solid line indicates null effect line. The scales are arranged in decreasing order of the RRs in trials deemed to be of high quality. Asterisk indicates unpublished scale.

Cochrane Criteria for RCTs

- Sequence generation
- Allocation concealment
- Blinding of participants, personnel and outcome assessors
- Incomplete outcome data
- Selective outcome reporting
- Other threats to validity

How to use Risk of Bias Assessment

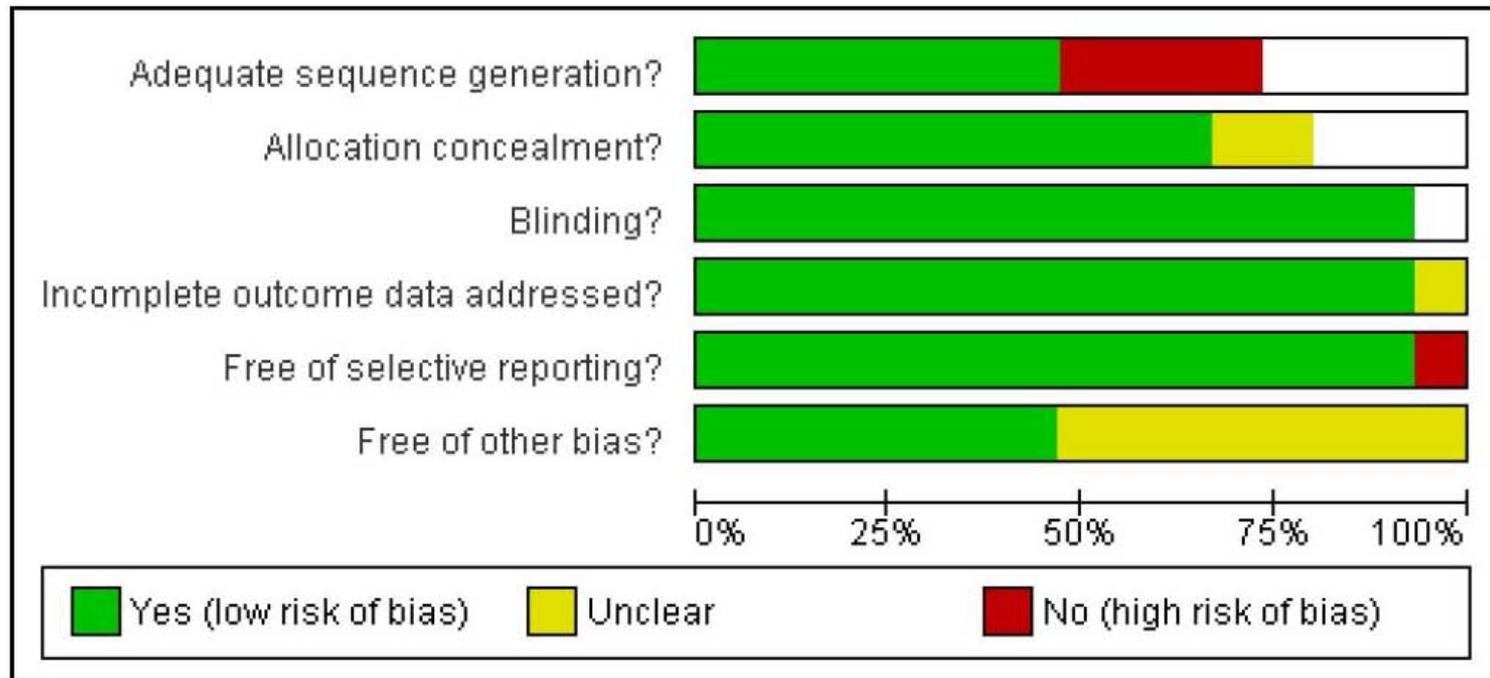
In meta-analysis

Exclude studies
Report descriptively
Sensitivity analyses

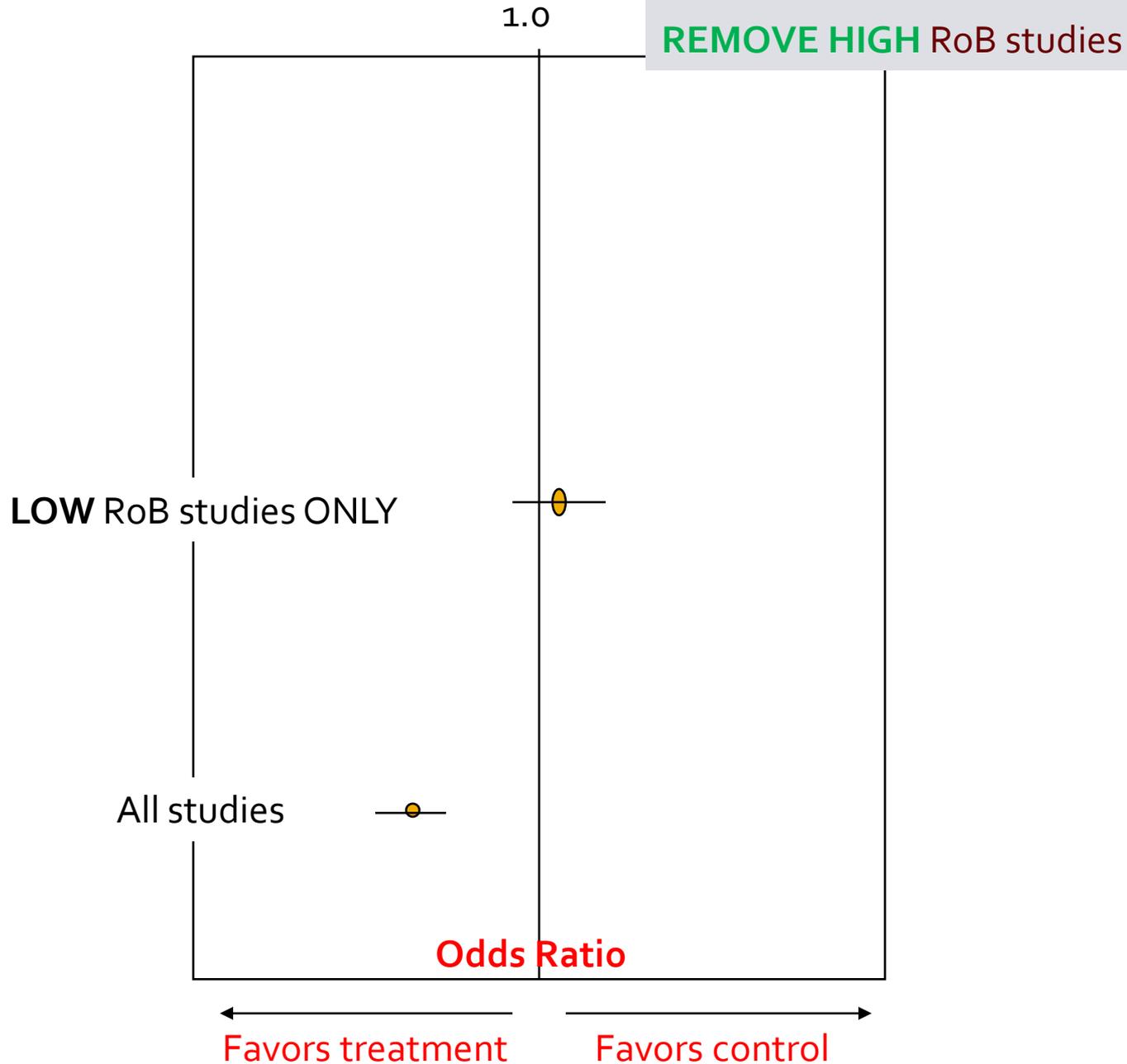
Figure 1. Methodological quality summary: review authors' judgements about each methodological quality item for each included study.

	Adequate sequence generation?	Allocation concealment?	Blinding?	Incomplete outcome data addressed?	Free of selective reporting?	Free of other bias?
Al-Nakib 1987	●	?	+	?	●	?
Douglas 1987	+	+	+	+	+	?
Farr 1987a		+	+	+	+	?
Godfrey 1992	+	+	+	+	+	?
Kurugol 2006a	+	+	+	+	+	+
Kurugol 2006b	+	+	+	+	+	+
Kurugol 2007	+	+	+	+	+	+
Macknin 1998	+	+	+	+	+	?
Mossad 1996	+	+	+	+	+	+
Petrus 1998	●		+	+	+	?
Prasad 2000		+	+	+	+	+
Prasad 2008		+	+	+	+	+
Smith 1989			+	+	+	?
Vakili 2009	●			+	+	+
Weismann 1990	●	?	+	+	+	?

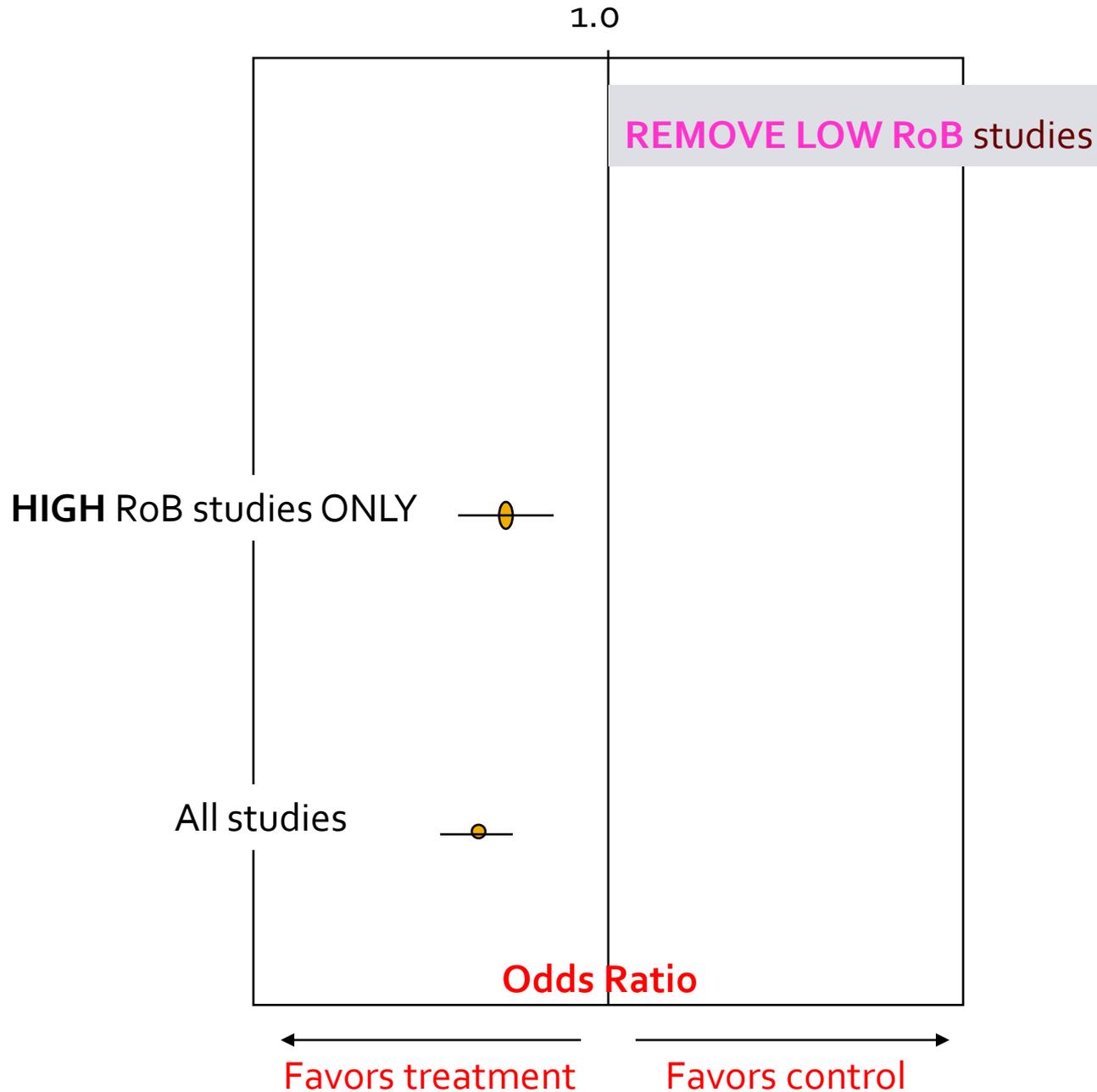
Figure 2. Methodological quality graph: review authors' judgements about each methodological quality item presented as percentages across all included studies.



Sensitivity analysis



Sensitivity analysis



Reporting

Reporting of clinical research has improved as risk of bias assessments for systematic reviews and other purposes became more prevalent and standards for reporting were implemented by journals.

Special Communication

Improving the Quality of Reporting of Randomized Controlled Trials

The CONSORT Statement

Colin Begg, PhD; Mildred Cho, PhD; Susan Eastwood, ELS(D); Richard Horton, MB; David Moher, MSc; Ingram Olkin, PhD; Roy Pitkin, MD; Drummond Rennie, MD; Kenneth F. Schulz, PhD; David Simel, MD; Donna F. Stroup, PhD

THE RANDOMIZED controlled trial (RCT), more than any other methodology, can have a powerful and immediate impact on patient care. Ideally, the report of such an evaluation needs to convey to the reader relevant information concerning the design, conduct, analysis, and generalizability of the trial. This information should provide the reader with the ability to make informed judgments regarding the internal and external validity of the trial. Accurate and complete reporting also benefits editors and reviewers in their deliberations regarding submitted manuscripts. For RCTs to ultimately benefit patients, the published report should be of the highest possible standard.

For editorial comment see p 649.

Evidence produced repeatedly over the last 30 years indicates a wide chasm between what a trial should report and what is actually published in the literature. In a review of 71 RCTs with negative results published between 1960 and 1975, the authors reported that the vast majority of them had too few patients to observe moderate or large differences.¹ Twenty years later, THE JOURNAL re-

ported research indicating few improvements in this situation and expressed a concern about the reporting of RCTs in general.²

In an effort to correct these and other problems, the Standards of Reporting Trials (SORT) group met on October 7 and 8, 1993. At the conclusion of the 2-day workshop, the SORT group put forth a new proposal for the reporting of RCTs: structured reporting.³ The proposal set out 24 essential items that needed to be included in the report of a trial, provided empirical evidence as to why the items should be included, and provided a format showing how they could be included.

Independently, approximately 5 months later (March 14 to 16, 1994), another group, the Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature, met to discuss similar challenges facing the reporting of clinical trials. Their proposal⁴ consisted of a checklist of items that should be included when reporting a clinical trial, along with a suggestion that editors add it to the Instructions for Authors.

A subsequent Editorial⁵ urged both groups to meet and decide which recommendations from each group's proposal should be retained. Besides being pragmatic, this suggestion had the potential for increasing consensus, which in turn might afford a greater chance of improving the quality of reporting of clinical trials to a wider audience.

On September 20, 1995, a total of 9 members (including editors, clinical epidemiologists, and statisticians) of the SORT group and the Asilomar Working Group met in Chicago, Ill. Two other people participated in the meeting: a journal editor (R.H.) who had expressed interest in helping to improve the reporting of RCTs and one of the authors (D.S.) of a trial report that used the SORT approach.⁶

METHODS

We started the day by reviewing both the SORT and Asilomar checklists to ascertain which items covered similar content areas and which ones were unique. Those items having similar content areas we then reviewed individually. We decided, a priori, to keep only those items for which there was empirical evidence, when available, that not reporting them resulted in bias in the estimates of the effects of interventions. We used common sense for those items included for which there was no empirical evidence. The selection of items was achieved using a modified Delphi process. We also emphasized the need to keep the number of items to a minimum, while maintaining adequate standards of reporting RCTs. We used a similar approach in deciding which of the unique items should remain in the resulting checklist. The day ended with a discussion on the use of the flow diagram proposed by the SORT group and the format of a trial report. Within a week or so following the meeting, a draft report was circulated to the entire group for further refinement. This process was continued until we felt the report accurately represented what had gone on during the meeting.

RESULTS

This meeting resulted in the Consolidated Standards of Reporting Trials (CONSORT) statement—a checklist (Table) and a flow diagram (Figure). The checklist consists of 21 items that pertain mainly to the methods, results, and discussion of an RCT report and identify key pieces of information necessary to evaluate the internal and external validity of the report. We have included at least 1 reference for each item, when appropriate (Table). The flow diagram provides information about the progress of patients throughout a 2-group parallel-

From the Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY (Dr Begg); Center for Bioethics, University of Pennsylvania, Philadelphia (Dr Cho); Department of Neurological Surgery, University of California, San Francisco (Ms Eastwood); *The Lancet*, London, United Kingdom (Dr Horton); Departments of Medicine and Epidemiology and Community Health, University of Ottawa (Ontario) (Mr Moher); Department of Statistics, Stanford (Calif) University (Dr Olkin); *Obstetrics and Gynecology*, Los Angeles, Calif (Dr Pitkin); *JAMA*, Chicago, Ill (Dr Rennie); Centers for Disease Control and Prevention, Atlanta, Ga (Drs Schulz and Stroup); and Center for Health Services Research in Primary Care, Durham (NC) Veterans Affairs Medical Center (Dr Simel).

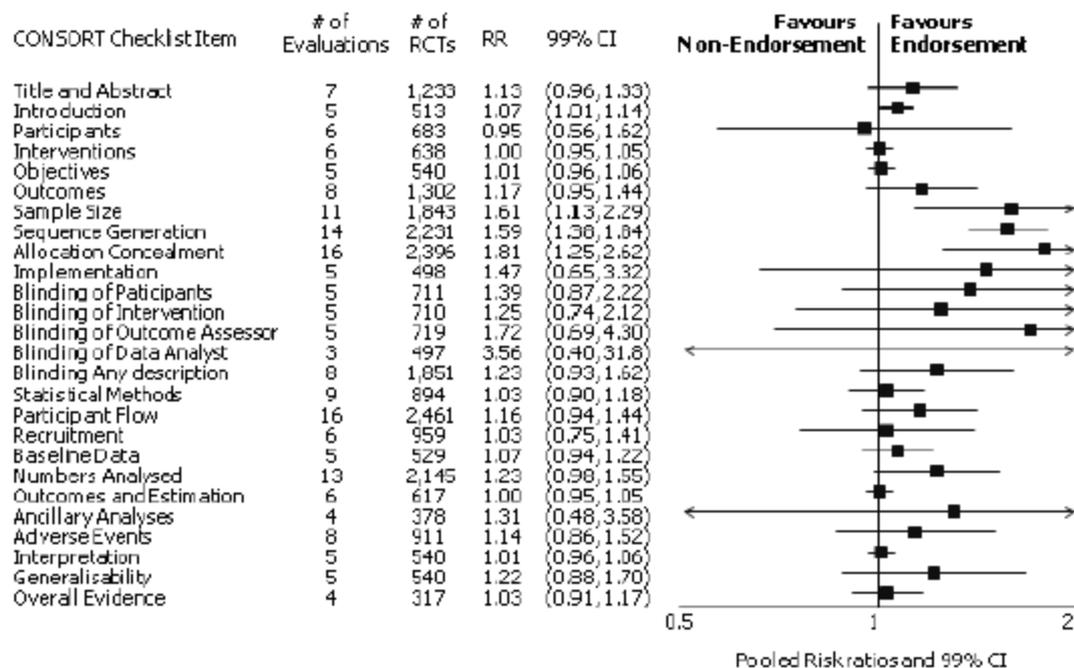
Reprints: David Moher, MSc, Clinical Epidemiology Unit, Loeb Medical Research Institute, Ottawa Civic Hospital, 1053 Carling Ave, Ottawa, Ontario, Canada K1Y 4E9 (e-mail: moher@ce.uottawa.ca).

Improvements in Reporting

- 50 evaluations of reporting in CONSORT-endorsing vs. Non-CONSORT endorsing journals
- 25 of 27 reporting elements improved post-CONSORT
- Largest improvements:
 - Allocation concealment (RR 1.81, CI 1.25, 2.61)
 - Sequence generation (RR 1.59, CI 1.38, 1.84)
 - Any mention of blinding (RR 1.23, CI 0.98, 1.55)
 - Sample size (1.61, CI 1.13, 2.29)

Improvements in Reporting

Figure 4. Pooled risk ratios across assessed 2001 CONSORT checklist items with 99% confidence intervals for primary comparison, adherence of RCTs published in CONSORT-endorsing journals versus RCTs published in CONSORT non-endorsing journals Plot generated in Comprehensive Meta-analysis Version 2.0 (CMA).



Citation: Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, Dias S, Schulz KF, Plint AC, Moher D. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews* 2012, Issue 11. Art. No.: MR000030. DOI: 10.1002/14651858.MR000030.pub2.

Bottom line on bias

- Assess specific risk of bias criteria relevant to the study design
- Don't confuse risk of bias with reporting criteria, imprecision
- Don't let a lack of reporting be an excuse for not assessing risk of bias
- Don't use "quality" or "methodology" scores
- Decide *a priori* what to do with the risk of bias assessment: exclude studies, report descriptively, sensitivity analysis