



# Quality of randomised trials in COPD

B. Bausch\*, A. Spaar\*, J. Kleijnen<sup>#,†</sup> and M.A. Puhan<sup>\*,†</sup>

**ABSTRACT:** Randomised trials can provide high-level evidence to inform treatment decisions. Since their quality in respiratory medicine is largely unknown, we assessed the quality of a large set of chronic obstructive pulmonary disease (COPD) trials.

As a marker of trial quality, we assessed the procedure and concealment of random allocation, and the conduct of an intention-to-treat-analysis in 344 randomised trials published between 1957 and 2006. We used ordered logistic regression to assess the association between trial quality and type of intervention, type of journal, journal impact factor and year of publication.

257 (75%) trials assessed pharmacological and 87 (25%) assessed nonpharmacological interventions. The generation of appropriate randomisation was reported in 27.0% of the trials, concealment of random allocation in 11.6% and an intention-to-treat analysis in 21.8% of trials. Significantly higher quality was found in trials on nonpharmacological interventions (OR 2.49, 95% CI 1.56–3.99), and in trials published in general medical journals (*versus* specialised journals; OR 2.25, 95% CI 1.30–3.90) and after 2000 (*versus* 1957–2000; OR 2.28, 95% CI 1.45–3.58). The association of quality with a high impact factor was of borderline significance ( $p=0.06$ ).

The quality of many COPD trials is low but tends to become better since the adoption of the CONSORT (Consolidated Standards of Reporting Trials) statement.

**KEYWORDS:** CONSORT (Consolidated Standards of Reporting Trials) statement, GRADE (Grades of Recommendation, Assessment, Development and Evaluation) system, quality of reporting, randomised trials, respiratory medicine

High-quality randomised controlled trials (RCTs) provide the most unbiased estimates of treatment effects and, ideally, support clinical practice to balance benefits and downsides of treatments. However, for physicians, it is often difficult to appraise the quality of evidence and to judge whether there are high-quality RCTs supporting the effectiveness of treatments. Evidence-based guidelines provide such support if they are based on systematic assessments of the available evidence. The American Thoracic Society and the European Respiratory Society, together with the GRADE working group (Grades of Recommendation, Assessment, Development and Evaluation), are currently undertaking significant efforts to develop evidence-based guidelines for chronic obstructive pulmonary disease (COPD), in which the grading of the methodological quality of evidence and the strength of recommendations are core components [1–3].

There is evidence from various medical fields that many RCTs suffer from important methodological limitations that are associated with biased results [4–10]. For example, if the method of

randomisation is not reported adequately, treatment effects tend to be overestimated. In addition, it is quite common practice to exclude patients from the analyses who do not adhere to the study treatment. Such “per-protocol” analysis is known to be associated with overestimation of treatment effects when compared with “intention-to-treat” analysis, the gold standard for the analysis of RCTs. In respiratory medicine, the quality of RCTs and factors associated with quality is largely unknown. Therefore, our aim was to analyse a large set of RCTs in COPD patients to determine the prevalence of high-quality RCTs and to explore which factors might influence their methodological quality.

## METHODS

### Selection criteria

As part of earlier systematic reviews on the effectiveness of long-term pharmacological (short- and long-acting bronchodilators, inhaled and systemic corticosteroids, mucolytics, antibiotics, immunostimulants, methylxanthines and almitrine, *etc.*) and nonpharmacological COPD treatments (respiratory rehabilitation, surgery

## AFFILIATIONS

\*Horten Centre for Patient-oriented Research and Knowledge Transfer, University of Zurich, Switzerland.

#Kleijnen Systematic Reviews Ltd, York, UK.

†Care and Public Health Research Institute (CAPHRI), University of Maastricht, Maastricht, The Netherlands.

\*Dept of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

## CORRESPONDENCE

M.A. Puhan  
615 North Wolfe Street  
Johns Hopkins Bloomberg School of Public Health  
W5010  
21205 Baltimore  
MD  
USA  
E-mail: mpuhan@jhsph.edu

## Received:

Oct 21 2008

Accepted after revision:

June 07 2009

First published online:

July 02 2009

and long-term oxygen therapy), we included randomised controlled trials with a follow-up of  $\geq 4$  weeks (with exception of antibiotics for exacerbation treatment). We accepted both a clinical diagnosis of COPD and spirometrically confirmed COPD since, in general, few trials published before 1995 reported results from spirometry. However, in order to detect all trials in which patients were very likely to have COPD, we included trials if the study populations had COPD, chronic bronchitis and emphysema, were aged above 40 yrs and if  $\geq 80\%$  were smokers or ex-smokers. The presence of these characteristics (age, chronic bronchitis and smoking history) renders the diagnosis of COPD very likely (positive likelihood ratio of 220) [11]. We excluded any studies of patients with acute bronchitis, pneumonia, asthma or bronchiectasis.

### Search strategy and identification of studies

From their inception to October 2006 the following databases were searched for relevant COPD studies using a very broad strategy: Medline, Embase, Latin American & Caribbean Health Science Literature, Cochrane Database of Systematic Reviews, The Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, Health Technology Assessment database, Cinahl, Biosis and the National Research Register. Keywords included *copd.mp.*, *coad.mp.*, *exp Lung Disease*, *Obstructive/*, *airflow obstruction*, *emphysema*, *hyperinflation*, *exp comparative study*, *exp double-blind method*, *exp single-blind method*, *controlled clinical trial* and *randomised controlled trial*, among others. We checked references of retrieved articles and we entered all included studies into related articles function of PubMed (last search for studies published after October 2006 on August 15, 2007). Finally, we used existing systematic reviews to identify further articles. The search strategies are available as online supplementary material.

### Study selection

Two members of the review team independently assessed the titles and abstracts of all identified citations without imposing any language restrictions. We ordered the full text of articles that seemed potentially eligible by one of the reviewers. Two reviewers then independently evaluated the full text of articles and decided upon inclusion and exclusion. If there was a disagreement about inclusion or exclusion after discussion, a third independent reviewer decided upon inclusion or exclusion.

### Quality assessment

Two reviewers independently assessed the quality of each included trial using a standardised form for RCTs [12]. We assessed the following markers of study quality: generation of the randomisation sequence, concealment of random allocation, description of eligibility criteria, comparability of study groups at baseline, masking of outcome assessors, treatment providers and patients and the performance of an intention-to-treat analysis. If scores were discordant based on real differences in interpretation, a third independent reviewer made a decision. Our primary outcome for the quality of trials included three important markers of internal validity that are known to be associated with substantial bias: procedure of random allocation, concealment of random allocation and the conduct of an intention-to-treat analysis [4]. We chose these three important aspects from all quality assessment criteria

because they apply to both pharmacological and nonpharmacological trials, whereas other aspects such as masking (blinding) of patients are mainly possible in pharmacological trials. Although masking of outcome assessors is possible in any trial, we did not include it in our primary analysis because many outcomes used in COPD trials (e.g. symptom scores or health-related quality of life) do not require an outcome assessor.

We considered the procedure of random allocation as adequate if the authors reported the use of computer-generated random numbers, tables of random numbers, tossing a coin or throwing dice. Concealment of random allocation was adequate if patients and investigators could not foresee group assignment, owing to, for example: centralised or pharmacy-controlled randomisation; serially numbered identical containers; on-site computer based system with a randomisation sequence not readable until allocation; or other approaches with robust methods to prevent foreknowledge of the allocation sequence to clinicians and patients [12]. Finally, we considered an intention-to-treat analysis to be carried out if the investigators described and presented it as such in the methods and results section of included papers. If the authors declared an analysis to be based on intention-to-treat (method section) but the number of patients finally included in the analyses differed from the number of randomised patients we did not consider it as an intention-to-treat analysis. We did not use a quality score such as the Jadad score, because there is general agreement that such scores do not adequately represent the extent of bias [13].

### Statistical analysis

We described the frequency of adequately reported aspects of internal validity using absolute numbers and proportions. Using simple and multiple ordered logistic regression analysis we assessed whether the number of adequately reported aspects of internal validity (0–3) was significantly associated with the type of intervention (nonpharmacological intervention *versus* pharmacological), the impact factor of the journal (impact factor  $\geq 5$ , 2–4.99 or  $<2$ ), the type of journal (general medicine *versus* respiratory medicine or other specialty journals) and the year of publication (after 2000, when the CONSORT (Consolidated Standards of Reporting Trials) statement was adopted by most journals, *versus* 1957–2000). We assessed the impact of the CONSORT statement because we hypothesised that it not only improved the reporting but also raised awareness for methodological aspects of trials and, therefore, may have improved the quality of trials. We tested the interaction between the years of publication and the type of intervention, type of journal and impact factor. Finally, we tested the parallel regression assumption of ordered logistic regression using the proportional odds test (Brant test) and did not find evidence that the parallel regression assumption has been violated. All analyses were conducted with STATA for Windows version 10.1 (Stata Corp., College Station, TX, USA).

### RESULTS

From 20,162 identified studies we included a total of 344 RCTs published between 1957 and 2006. 257 (75%) of the trials assessed pharmacological interventions: bronchodilators ( $n=61$ ), inhaled corticosteroids ( $n=29$ ), systemic corticosteroids

(n=10), mucolytics (n=43), antibiotics (n=14), immunostimulants (n=43), methylxanthines (n=6), almitrine (n=26) and others (n=25). 87 (25%) of the trials assessed nonpharmacological COPD treatments, including rehabilitation (n=60), smoking cessation (n=8), lung volume reduction surgery (n=10) and oxygen therapy (n=9) (fig. 1).

Description of quality assessment

Table 1 summarises the quality assessment of the 344 RCTs. The generation of appropriate randomisation was reported in 27.0% of the trials. Concealment of randomisation was reported in 11.6% and the conduct of an intention-to-treat analysis in 21.8% of the RCTs, respectively. Masking of patients and treatment providers was reported in a majority of trials, and in drug trials in particular. But masking of outcome assessors was reported substantially less frequently (20.6%). Eligibility criteria were reported adequately in most trials, although nonpharmacological trials tended to show better reporting (84.3%) than drug trials (64.2%). Finally, 28.5% reported confidence intervals around treatment effect estimates, whereas the majority of trials did not provide measures of precision.

Quality of trials over time

Figure 2 shows the proportion of COPD trials adequately reporting on important aspects of internal validity over time. Early trials (1957–1974) were of higher quality than later trials

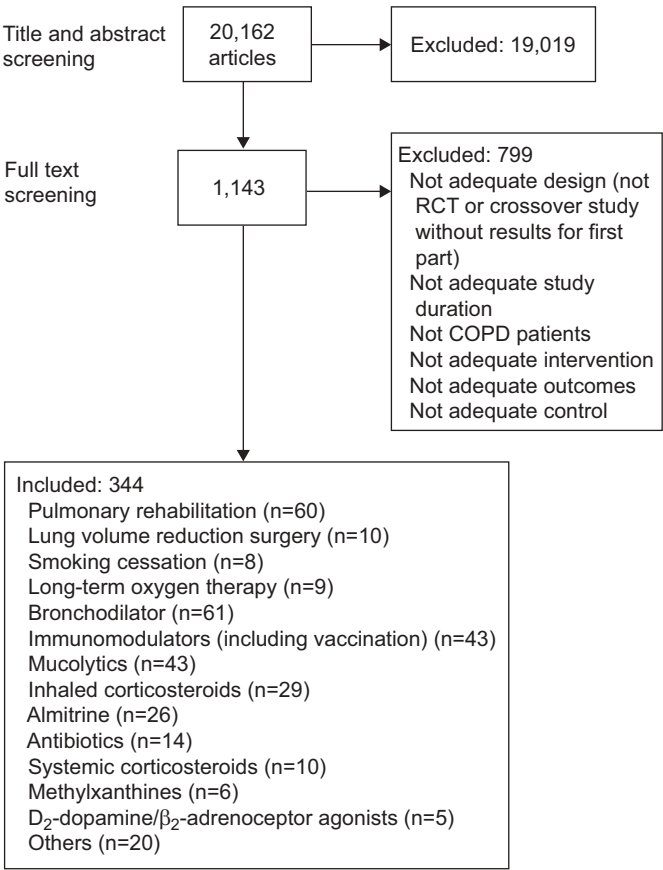


FIGURE 1. Study selection. RCT: randomised controlled trial; COPD: chronic obstructive pulmonary disease.

TABLE 1 Quality assessment of 344 randomised controlled trials in chronic obstructive pulmonary disease									
Reporting of Interventions	Subjects	Generation of randomisation	Concealment of random allocation	Intention-to-treat analysis	Masking of outcome assessor	Masking of patients	Masking of treatment provider	Eligibility criteria specified	Confidence intervals reported
All	344	93 (27)	40 (11.6)	75 (21.8)	71 (20.6)	182 (52.9)	177 (51.5)	238 (69.2)	98 (28.5)
Pharmacological treatment	257	57 (22.2)	28 (10.9)	49 (19.1)	53 (20.6)	175 (68.1)	171 (66.5)	165 (64.2)	64 (24.9)
Rehabilitation	60	24 (40)	5 (8.3)	14 (23.3)	11 (18.3)	NA	NA	50 (83.3)	25 (41.7)
Smoking cessation	8	3 (37.5)	2 (25)	5 (62.5)	5 (62.5)	3 (37.5)	3 (37.5)	6 (75)	3 (37.5)
Oxygen	9	4 (44.4)	1 (11.1)	4 (44.4)	NR	1 (11.1)	1 (11.1)	8 (88.9)	1 (11.1)
Surgery	10	5 (50)	4 (40)	3 (30)	2 (20)	NA	NA	9 (90)	5 (50)

Data are presented as n or n (%). NR: not reported; NA: not applicable.

but the number of trials from that period was small ( $n=11$ ) and 95% confidence intervals overlapped with those of later periods. There was a steady increase in the proportion of trials reporting on at least two aspects of internal validity from 1975 on but it never exceeded 20%.

#### **Association of trial quality with type of intervention, type of journal, impact factor and year of publication**

In the simple regression analysis, trials on nonpharmacological interventions were significantly more likely to be of higher quality than pharmacological trials (OR 2.41, 95% CI 1.52–3.81; table 2). Also, trials published in general medical journals, in journals with an impact factor  $\geq 5$  and trials published after the year 2000 were significantly more likely to be of higher methodological quality. In the multiple regression analysis type of intervention, type of journal and year of publication remained significant predictors of higher quality, whereas the association of an impact factor of  $\geq 5$  with trial quality was only of borderline significance. Finally, the interaction between years of publication (after 2000 *versus* 1957–2000) and type of journal (general medical *versus* specialised journals) was significant ( $p=0.05$ ) in multiple regression analyses, whereas interactions with type of intervention and impact factor were not.

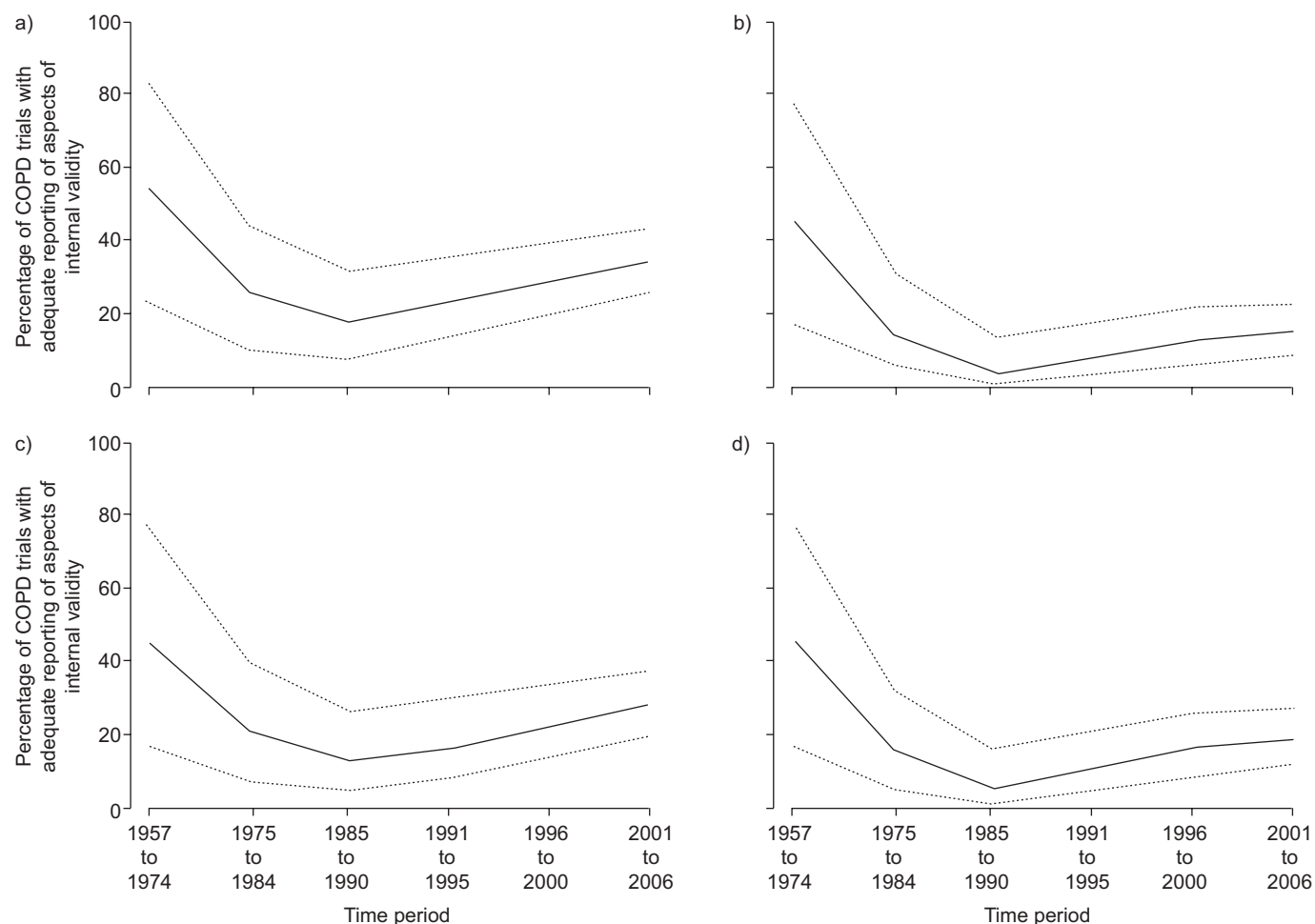
## **DISCUSSION**

### **Main findings**

This study showed that in a large set of pharmacological and nonpharmacological COPD trials the majority of trials was of low methodological quality with only every fifth trial reporting adequately on the generation of randomisation, concealment of random allocation or an intention-to-treat analysis. These methodological limitations are known to substantially bias treatment effects towards an overestimation [4–8, 10]. Trials on nonpharmacological interventions, trials published in general medical journals and trials published after the year 2000 were significantly more likely to be of higher methodological quality.

### **Strengths and limitations**

The strengths of this study are the large number of reviewed COPD trials, which covered a broad therapeutic spectrum from pharmacological to nonpharmacological interventions. The extensive literature search and the wide inclusion criteria led to a set of RCTs that is representative of COPD trials and, perhaps, also of trials in respiratory medicine in general. In addition, two reviewers independently assessed three important aspects of internal validity.



**FIGURE 2.** Proportion (95% CI) of chronic obstructive pulmonary disease (COPD) trials adequately reporting on a) generation of randomisation, b) concealment of random allocation, c) intention-to-treat analysis over time and d) two or more aspects of internal validity.



**TABLE 2** Association of type of intervention, type of journal, impact factor and year of publication with quality of randomised controlled trials in chronic obstructive pulmonary disease

	OR per one additional aspect of internal validity <sup>#</sup>			
	Univariate associations		Associations from multiple regression analysis	
	OR (95% CI)	p-value	OR (95% CI)	p-value
<b>Type of intervention</b>				
Nonpharmacological (n=87) versus pharmacological (n=257) treatments	2.41 (1.52–3.81)	p<0.001	2.49 (1.56–3.99)	p<0.001
<b>Journal type</b>				
General medical (n=75) versus specialised (n=269) journals	1.70 (1.02–2.82)	p=0.04	2.25 (1.30–3.90)	p=0.004
<b>Impact factor</b>				
2–4.99 (n=97) versus <2 (n=125)	1.54 (0.91–2.61)	p=0.11	1.41 (0.82–2.44)	p=0.22
≥5 (n=122) versus <2 (n=125)	1.98 (1.20–3.29)	p=0.008	1.65 (0.98–2.78)	p=0.06
<b>Year of publication<sup>*</sup></b>				
2001–2006 versus 1957–2000	1.95 (1.27–2.99)	p=0.002	2.28 (1.45–3.58)	p<0.001

<sup>#</sup>: generation of random sequence, concealment of random allocation and intention-to-treat analysis. <sup>\*</sup>: before/after introduction of CONSORT (Consolidated Standards of Reporting Trials) statement.

A limitation of our study is that we could only assess the reporting of important markers of internal validity but not what the investigators actually did. Although the reporting and the actual quality of trials correlate it is also possible that some misclassification occurred [14–17]. Furthermore, we limited the main analysis to three aspects of internal validity and did not consider other important sources of bias, such as masking of patients.

### Reporting of RCTs

Low-quality trials are associated with overestimated treatment estimates [4–8]. The CONSORT statement was developed and revised in order to improve RCTs and guarantee that they were of high quality [18, 19]. As our analysis showed, the quality of trials published after 2000 was higher compared to earlier trials, a difference that was most pronounced in general medical journals. This is not surprising given that journals such as *Annals of Internal Medicine*, *British Medical Journal*, *Lancet* and *Journal of the American Medical Association* were the driving forces behind the efforts to improve the reporting of trials. It is, however, unclear whether this finding is entirely due to the CONSORT statement that was adopted around the year 2000 by many major journals. Our analysis shows that there was a slow but steady increase of trial quality over time, without a pronounced increase around or after the year 2000.

Although the CONSORT statement provides primarily guidance for the reporting of RCTs it is also a valuable resource for investigators during the planning stage of a trial [18, 19]. The extensive background document and its references provide ample information on how to optimally design a RCT. Despite the development of a transparent and easily understandable CONSORT statement and its wide dissemination in almost all major medical journals, our study shows that the quality of RCTs in COPD is still low. Similar studies also found that the reporting of key methodological items continued to be poor, even in the major general medical journals [20]. It is not evident to us why

the quality of pharmacological trials is lower compared to nonpharmacological trials. One reason could be that the design and conduct of trials on nonpharmacological interventions is more complex than for pharmacological trials. Such studies may require a more careful and longer planning stage, which may be reflected in their higher methodological quality.

The low quality of trials is not only of concern to clinicians who may be willing to adopt the results from RCTs for their own practice. It also complicates current efforts of the American Thoracic Society and the European Respiratory Society to develop evidence-based guidelines for COPD [2, 3]. It is more difficult to provide unambiguous recommendations for clinical practice if the underlying evidence is of low quality.

What can be done to improve the quality of RCTs in COPD research or respiratory medicine in general? The awareness of authors and editors of the CONSORT statement should continue to be raised. There is no evidence regarding the awareness of the CONSORT statement in the respiratory medicine community but our results imply that it is rather low. If the CONSORT statement and its background document were considered more seriously by authors and editors it is likely that not only the reporting of RCTs would improve but also that important aspects of internal validity are considered to a larger extent during the planning stage of a trial.

### Conclusion

This study showed that the quality of RCTs in COPD is low. Even high-impact journals have often published RCTs that are at substantial risk for bias. Editors should continue to encourage authors to adhere strongly to the CONSORT statement when reporting their RCTs or, even better, emphasise the need to consider important aspects of internal validity during the planning stage of a trial. Thereby, RCTs will provide more valid estimates of treatment effects and serve as a reliable basis for the development of evidence-based guidelines.

## SUPPORT STATEMENT

M.A. Puhan is supported by a career award of the Swiss National Science Foundation (#3233B0/115216/1).

## STATEMENT OF INTEREST

Statements of interest for all authors can be found at [www.erj.ersjournals.com/misc/statements.dtl](http://www.erj.ersjournals.com/misc/statements.dtl)

## REFERENCES

- 1 Guyatt G, Gutterman D, Baumann MH, *et al.* Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians task force. *Chest* 2006; 129: 174–181.
- 2 Schunemann HJ, Jaeschke R, Cook DJ, *et al.* An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006; 174: 605–614.
- 3 Schunemann HJ, Woodhead M, Anzueto A, *et al.* A vision statement on guideline development for respiratory disease: the example of COPD. *Lancet* 2009; 373: 774–779.
- 4 Schulz KF, Chalmers I, Hayes RJ, *et al.* Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408–412.
- 5 Moher D, Pham B, Jones A, *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352: 609–613.
- 6 DerSimonian R, Charette LJ, McPeck B, *et al.* Reporting on methods in clinical trials. *N Engl J Med* 1982; 306: 1332–1337.
- 7 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987; 317: 426–432.
- 8 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; 272: 122–124.
- 9 Siersma V, Als-Nielsen B, Chen W, *et al.* Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med* 2007; 26: 2745–2758.
- 10 Wood L, Egger M, Gluud LL, *et al.* Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336: 601–605.
- 11 Straus SE, McAlister FA, Sackett DL, *et al.* The accuracy of patient history, wheezing, and laryngeal measurements in diagnosing obstructive airway disease. CARE-COAD1 Group. Clinical Assessment of the Reliability of the Examination-Chronic Obstructive Airways Disease. *JAMA* 2000; 283: 1853–1857.
- 12 Khan KS, ter Riet G, Glanville J, *et al.*, Undertaking Systematic Reviews of Research on Effectiveness. 2nd Edn. York, Centre for Reviews and Dissemination, 2001.
- 13 Juni P, Witschi A, Bloch R, *et al.* The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282: 1054–1060.
- 14 Soares HP, Daniels S, Kumar A, *et al.* Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004; 328: 22–24.
- 15 Devereaux PJ, Choi PT, El-Dika S, *et al.* An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *J Clin Epidemiol* 2004; 57: 1232–1236.
- 16 Pildal J, Chan AW, Hrobjartsson A, *et al.* Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ* 2005; 330: 1049.
- 17 Huwiler-Muntener K, Juni P, Junker C, *et al.* Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA* 2002; 287: 2801–2804.
- 18 Begg C, Cho M, Eastwood S, *et al.* Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276: 637–639.
- 19 Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; 285: 1987–1991.
- 20 Mills EJ, Wu P, Gagnier J, Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. *Contemp Clin Trials* 2005; 26: 480–487.