

# The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration

Douglas G. Altman, DSc; Kenneth F. Schulz, PhD; David Moher, MSc; Matthias Egger, MD; Frank Davidoff, MD; Diana Elbourne, PhD; Peter C. Gøtzsche, MD; and Thomas Lang, MA, for the CONSORT Group

Overwhelming evidence now indicates that the quality of reporting of randomized, controlled trials (RCTs) is less than optimal. Recent methodologic analyses indicate that inadequate reporting and design are associated with biased estimates of treatment effects. Such systematic error is seriously damaging to RCTs, which boast the elimination of systematic error as their primary hallmark. Systematic error in RCTs reflects poor science, and poor science threatens proper ethical standards.

A group of scientists and editors developed the CONSORT (Consolidated Standards of Reporting Trials) statement to improve the quality of reporting of RCTs. The statement consists of a checklist and flow diagram that authors can use for reporting an RCT. Many leading medical journals and major international editorial groups have adopted the CONSORT statement. The CONSORT statement facilitates critical appraisal and interpretation of RCTs by providing guidance to authors about how to improve the

reporting of their trials.

This explanatory and elaboration document is intended to enhance the use, understanding, and dissemination of the CONSORT statement. The meaning and rationale for each checklist item are presented. For most items, at least one published example of good reporting and, where possible, references to relevant empirical studies are provided. Several examples of flow diagrams are included.

The CONSORT statement, this explanatory and elaboration document, and the associated Web site (<http://www.consort-statement.org>) should be helpful resources to improve reporting of randomized trials.

*Ann Intern Med.* 2001;134:663-694.

[www.annals.org](http://www.annals.org)

For author affiliations and current addresses, see end of text.

The RCT is a very beautiful technique, of wide applicability, but as with everything else there are snags. When humans have to make observations there is always the possibility of bias (1).

**W**ell-designed and properly executed randomized, controlled trials (RCTs) provide the best evidence on the efficacy of health care interventions\*, but trials with inadequate methodologic approaches are associated with exaggerated treatment effects (2–5). Biased\* results from poorly designed and reported trials can mislead decision making in health care at all levels, from treatment decisions for the individual patient to formulation of national public health policies.

Critical appraisal of the quality of clinical trials is possible only if the design, conduct, and analysis of RCTs are thoroughly and accurately described in published articles. Far from being transparent, the reporting of RCTs is often incomplete (6–9), compounding problems arising from poor methodology (10–15).

## INCOMPLETE AND INACCURATE REPORTING

Many reviews have documented deficiencies in reports of clinical trials. For example, information on

whether assessment of outcomes\* was blinded was reported in only 30% of 67 trial reports in four leading journals in 1979 and 1980 (16). Similarly, only 27% of 45 reports published in 1985 defined a primary end point\* (14), and only 43% of 37 trials with negative findings published in 1990 reported a sample size\* calculation (17). Reporting is not only frequently incomplete but also sometimes inaccurate. Of 119 reports stating that all participants\* were included in the analysis in the groups to which they were originally assigned (intention-to-treat\* analysis), 15 (13%) excluded patients or did not analyze all patients as allocated (18). Many other reviews have found that inadequate reporting was common in specialty journals (19–29) and journals published in languages other than English (30, 31).

Proper randomization\* eliminates selection bias\* and is the crucial component of high-quality RCTs (32). Successful randomization hinges on two steps: generation\* of an unpredictable allocation sequence and concealment\* of this sequence from the investigators enrolling participants (Table 1) (2, 21). Unfortunately, reporting of the methods used for allocation of participants to interventions is also generally inadequate. For

Throughout the text, terms marked with an asterisk are defined at end of text.

**Table 1. Treatment Allocation. What's So Special about Randomization?**

The method used to assign treatments or other interventions to trial participants is a crucial aspect of clinical trial design. Random assignment\* is the preferred method; it has been successfully used in trials for more than 50 years (33). Randomization has three major advantages (34). First, it eliminates bias in the assignment of treatments. Without randomization, treatment comparisons may be prejudiced, whether consciously or not, by selection of participants of a particular kind to receive a particular treatment. Second, random allocation facilitates blinding\* the identity of treatments to the investigators, participants, and evaluators, possibly by use of a placebo, which reduces bias after assignment of treatments (35). Third, random assignment permits the use of probability theory to express the likelihood that any difference in outcome\* between intervention groups merely reflects chance (36). Preventing selection and confounding\* biases is the most important advantage of randomization (37).

Successful randomization in practice depends on two interrelated aspects: adequate generation of an unpredictable allocation sequence and concealment of that sequence until assignment occurs (2, 21). A key issue is whether the schedule is known or predictable by the people involved in allocating participants to the comparison groups\* (38). The treatment allocation system should thus be set up so that the person enrolling participants does not know in advance which treatment the next person will get, a process termed *allocation concealment\** (2, 21). Proper allocation concealment shields knowledge of forthcoming assignments, whereas proper random sequences prevent correct anticipation of future assignments based on knowledge of past assignments.

Terms marked with an asterisk are defined in the glossary at the end of the text.

example, at least 5% of 206 reports of supposed RCTs in obstetrics and gynecology journals described studies that were not truly randomized (21). This estimate is conservative, as most reports do not at present provide adequate information about the method of allocation (19, 21, 23, 25, 30, 39).

### IMPROVING THE REPORTING OF RCTs: THE CONSORT STATEMENT

DerSimonian and colleagues (16) suggested that “editors could greatly improve the reporting of clinical trials by providing authors with a list of items that they expected to be strictly reported.” Early in the 1990s, two groups of journal editors, trialists, and methodologists independently published recommendations on the reporting of trials (40, 41). In a subsequent editorial, Rennie (42) urged the two groups to meet and develop a common set of recommendations; the outcome was the CONSORT statement (*Consolidated Standards of Reporting Trials*) (43).

The CONSORT statement (or simply CONSORT) comprises a checklist of essential items that should be included in reports of RCTs and a diagram for documenting the flow of participants through a trial. It is aimed at first reports of two-group parallel designs.

Most of CONSORT is also relevant to a wider class of trial designs, such as equivalence, factorial, cluster, and crossover trials. Modifications to the CONSORT checklist for reporting trials with these and other designs are in preparation.

The objective of CONSORT is to facilitate critical appraisal and interpretation of RCTs by providing guidance to authors about how to improve the reporting of their trials. Peer reviewers and editors can also use CONSORT to help them identify reports that are difficult to interpret and those with potentially biased results. However, CONSORT was not meant to be used as a quality assessment instrument. Rather, the content of CONSORT focuses on items related to the internal and external validity\* of trials. Many items not explicitly mentioned in CONSORT should also be included in a report, such as information about approval by an ethics committee, obtaining of informed consent from participants, existence of a data safety and monitoring committee, and sources of funding. In addition, other aspects of a trial should be properly reported, such as information pertinent to cost-effectiveness analysis (44–46) and quality-of-life assessments (47).

### THE REVISED CONSORT STATEMENT: EXPLANATION AND ELABORATION

Since its publication in 1996, CONSORT has been supported by an increasing number of journals (48–51) and several editorial groups, including the International Committee of Medical Journal Editors (the Vancouver Group) (52). Evidence is accumulating that the introduction of CONSORT has improved the quality of reports of RCTs (53, 54). However, CONSORT is an ongoing initiative, and the statement is revised periodically (3). The 1996 version of the statement (43) received much comment and some criticism. For example, Meinert (55) pointed out that the terminology used lacked clarity and that the information presented in the flow diagram was incomplete. Work on a revised statement started in 1999; the revised checklist is shown in Table 2 and the revised flow diagram in Figure 1 (56–58).

During revision, it became clear that explanation and elaboration of the principles underlying the CONSORT statement would help investigators and others to write or appraise trial reports. In this article, we discuss the rationale and scientific background for each item

**Table 2. Checklist of Items To Include When Reporting a Randomized Trial†**

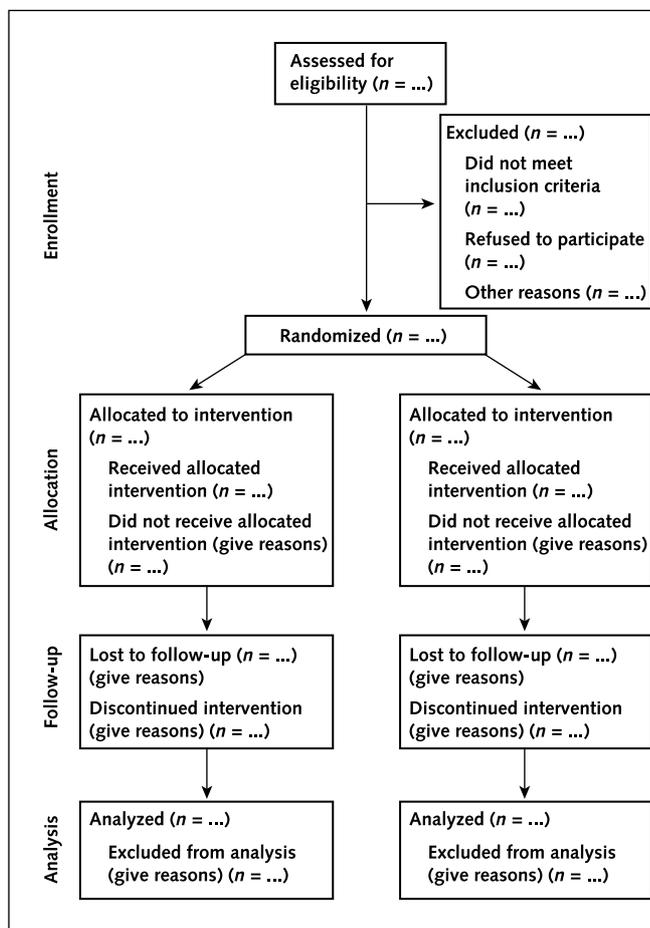
Paper Section and Topic	Item Number	Descriptor	Reported on Page Number
Title and abstract	1	How participants were allocated to interventions (e.g., "random allocation," "randomized," or "randomly assigned").	
Introduction Background	2	Scientific background and explanation of rationale.	
Methods			
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	
Objectives	5	Specific objectives and hypotheses.	
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	
Randomization			
Sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification).	
Allocation concealment	9	Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.	
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses.	
Results			
Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	
Recruitment	14	Dates defining the periods of recruitment and follow-up.	
Baseline data	15	Baseline demographic and clinical characteristics of each group.	
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat." State the results in absolute numbers when feasible (e.g., 10 of 20, not 50%).	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g., 95% confidence interval).	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	
Discussion			
Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes.	
Generalizability	21	Generalizability (external validity) of the trial findings.	
Overall evidence	22	General interpretation of the results in the context of current evidence.	

† From references 56–58.

(Table 2) and provide published examples of good reporting. (For further examples, see [www.consort-statement.org](http://www.consort-statement.org)). In these examples, we have removed authors' references to other publications to avoid confusion;

however, relevant references should always be cited where needed, such as to support unfamiliar methodologic approaches. Where possible, we describe the findings of relevant empirical studies. Many excellent books

**Figure 1. Revised template of the CONSORT (Consolidated Standards of Reporting Trials) diagram showing the flow of participants through each stage of a randomized trial (56–58).**



on clinical trials offer fuller discussion of methodologic issues (59–61).

For convenience, we sometimes refer to “treatments” and “patients,” although we recognize that not all interventions evaluated in RCTs are technically treatments and the participants in trials are not always patients.

**CHECKLIST ITEMS**

**Title and Abstract**

*Item 1.* How participants were allocated to interventions (e.g., “random allocation,” “randomized,” or “randomly assigned”).

*Examples*

Title: “Smoking reduction with oral nicotine inhalers: double blind, randomised clinical trial of efficacy and safety” (62).

Abstract: “Design: Randomized, double-blind, placebo-controlled trial” (63).

*Explanation*

The ability to identify a relevant report in an electronic database depends to a large extent on how it was indexed. Indexers for the National Library of Medicine’s MEDLINE database may not classify a report as an RCT if the authors do not explicitly report this information. To help ensure that a study is appropriately indexed as an RCT, authors should state explicitly in the abstract of their report that the participants were randomly assigned to the comparison groups. Possible wordings include “participants were randomly assigned to . . .,” “treatment was randomized,” or “participants were assigned to interventions by using random allocation.” We also strongly encourage the use of the word “randomized” in the title of the report to permit instant identification.

In the mid-1990s, electronic searching of MEDLINE yielded only about half of all RCTs relevant to a topic (64). This deficiency has been remedied in part by the work of the Cochrane Collaboration, which by 1999 had identified almost 100 000 RCTs that had not been indexed as such in MEDLINE. These reports have been reindexed (65). Adherence to this recommendation should improve the accuracy of indexing in the future.

We encourage the use of structured abstracts when a summary of the report is required. Structured abstracts provide readers with a series of headings pertaining to the design, conduct, and analysis of a trial; standardized information appears under each heading (66). Some studies have found that structured abstracts are of higher quality than the more traditional descriptive abstracts (67) and that they allow readers to find information more easily (68).

**Introduction**

*Item 2.* Scientific background and explanation of rationale.

### Example

The carpal tunnel syndrome is caused by compression of the median nerve at the wrist and is a common cause of pain in the arm, particularly in women. Injection with corticosteroids is one of the many recommended treatments.

One of the techniques for such injection entails injection just proximal to (not into) the carpal tunnel. The rationale for this injection site is that there is often a swelling at the volar side of the forearm, close to the carpal tunnel, which might contribute to compression of the median nerve. Moreover, the risk of damaging the median nerve by injection at this site is lower than by injection into the narrow carpal tunnel. The rationale for using lignocaine (lidocaine) together with corticosteroids is twofold: the injection is painless, and diminished sensation afterwards shows that the injection was properly carried out.

We investigated in a double blind randomised trial, firstly, whether symptoms disappeared after injection with corticosteroids proximal to the carpal tunnel and, secondly, how many patients remained free of symptoms at follow up after this treatment (69).

### Explanation

Typically, the introduction consists of free-flowing text, without a structured format, in which authors explain the scientific background or context and the scientific rationale for their trial. The rationale may be explanatory (for example, to compare the bioavailability of two formulations of a drug or assess the possible influence of a drug on renal function) or pragmatic (for example, to guide practice by comparing the clinical effects of two alternative treatments). Authors should report the evidence of the benefits of any active intervention included in a trial. They should also suggest a plausible explanation for how the intervention under investigation might work, especially if there is little or no previous experience with the intervention (70).

The Helsinki Declaration states that biomedical research involving people should be based on a thorough knowledge of the scientific literature (71). That is, it is unethical to expose human subjects unnecessarily to the risks of research. Some clinical trials have been shown to have been unnecessary because the question they addressed had been or could have been answered by a systematic review of the existing literature (72). Thus, the need for a new trial should be justified in the intro-

duction. Ideally, the introduction should include a reference to a systematic review of previous similar trials or a note of the absence of such trials (73).

In the first part of the introduction, authors should describe the problem that necessitated the work. The nature, scope, and severity of the problem should provide the background and a compelling rationale for the study. This information is often missing from reports. Authors should then describe briefly the broad approach taken to studying the problem. It may also be appropriate to include here the objectives\* of the trial (item 5).

### Methods

#### *Item 3a.* Eligibility criteria for participants.

### Example

... all women requesting an IUCD [intrauterine contraceptive device] at the Family Welfare Centre, Kenyatta National Hospital, who were menstruating regularly and who were between 20 and 44 years of age, were candidates for inclusion in the study. They were not admitted to the study if any of the following criteria were present: (1) a history of ectopic pregnancy, (2) pregnancy within the past 42 days, (3) leiomyomata of the uterus, (4) active PID [pelvic inflammatory disease], (5) a cervical or endometrial malignancy, (6) a known hypersensitivity to tetracyclines, (7) use of any antibiotics within the past 14 days or long-acting injectable penicillin, (8) an impaired response to infection, or (9) residence outside the city of Nairobi, insufficient address for follow-up, or unwillingness to return for follow-up (74).

### Explanation

Every RCT addresses an issue relevant to some population with the condition of interest. Trialists usually restrict this population by using eligibility criteria\* and by performing the trial in one or a few centers. Typical selection criteria may relate to age, sex, clinical diagnosis, and comorbid conditions; exclusion criteria are often used to ensure patient safety. Eligibility criteria should be explicitly defined. If relevant, any known inaccuracy in patients' diagnoses should be discussed because it can affect the power\* of the trial (75). The common distinction between inclusion and exclusion criteria is unnecessary (76).

Careful descriptions of the trial participants and the

setting in which they were studied are needed so that readers may assess the external validity (generalizability) of the trial results (item 21). Of particular importance is the method of recruitment\*, such as by referral or self-selection (for example, through advertisements). Because they are applied before randomization, eligibility criteria do not affect the internal validity of a trial, but they do affect the external validity.

Despite their importance, eligibility criteria are often not reported adequately. For example, 25% of 364 reports of RCTs in surgery did not specify the eligibility criteria (77). Eight published trials leading to clinical alerts by the National Institutes of Health specified an average of 31 eligibility criteria. Only 63% of the criteria were mentioned in the journal articles, and only 19% were mentioned in the clinical alerts (78). The number of eligibility criteria in cancer trials increased markedly between the 1970s and 1990s (76).

*Item 3b.* The settings and locations where the data were collected.

#### Example

Volunteers were recruited in London from four general practices and the ear, nose, and throat outpatient department of Northwick Park Hospital. The prescribers were familiar with homoeopathic principles but were not experienced in homoeopathic immunotherapy (79).

#### Explanation

Settings and locations affect the external validity of a trial. Health care institutions vary greatly in their organization, experience, and resources and the baseline risk for the medical condition under investigation. Climate and other physical factors, economics, geography, and the social and cultural milieu can all affect a study's external validity.

Authors should report the number and type of settings and care providers involved so that readers can assess external validity. They should describe the settings and locations in which the study was carried out, including the country, city, and immediate environment (for example, community, office practice, hospital clinic, or inpatient unit). In particular, it should be clear whether the trial was carried out in one or several centers ("multicenter trials"). This description should provide enough information that readers can judge whether the results of

the trial are relevant to their own setting. Authors should also report any other information about the settings and locations that could influence the observed results, such as problems with transportation that might have affected patient participation.

*Item 4.* Precise details of the interventions intended for each group and how and when they were actually administered.

#### Example

Patients with psoriatic arthritis were randomised to receive either placebo or etanercept (Enbrel) at a dose of 25 mg twice weekly by subcutaneous administration for 12 weeks . . . Etanercept was supplied as a sterile, lyophilised powder in vials containing 25 mg etanercept, 40 mg mannitol, 10 mg sucrose, and 1–2 mg tromethamine per vial. Placebo was identically supplied and formulated except that it contained no etanercept. Each vial was reconstituted with 1 mL bacteriostatic water for injection (80).

#### Explanation

Authors should describe each intervention thoroughly, including control interventions. The characteristics of a placebo and the way in which it was disguised should also be reported. It is especially important to describe thoroughly the "usual care" given to a control group or an intervention that is in fact a combination of interventions.

In some cases, description of who administered treatments is critical because it may form part of the intervention. For example, with surgical interventions, it may be necessary to describe the number, training, and experience of surgeons in addition to the surgical procedure itself (81).

When relevant, authors should report details of the timing and duration of interventions, especially if multiple-component interventions were given.

*Item 5.* Specific objectives and hypotheses.

#### Example

In the current study we tested the hypothesis that a policy of active management of nulliparous labour would: 1. reduce the rate of caesarean section, 2. reduce the rate of prolonged labour; 3. not influence maternal satisfaction with the birth experience (82).

### Explanation

Objectives are the questions that the trial was designed to answer. They often relate to the efficacy of a particular therapeutic or preventive intervention. Hypotheses\* are prespecified questions being tested to help meet the objectives.

Hypotheses are more specific than objectives and are amenable to explicit statistical evaluation. In practice, objectives and hypotheses are not always easily differentiated, as in the example above.

Some evidence suggests that the majority of reports of RCTs provide adequate information about trial objectives and hypotheses (24).

*Item 6a.* Clearly defined primary and secondary outcome measures.

### Example

The primary endpoint with respect to efficacy in psoriasis was the proportion of patients achieving a 75% improvement in psoriasis activity from baseline to 12 weeks as measured by the PASI [psoriasis area and severity index]. Additional analyses were done on the percentage change in PASI scores and improvement in target psoriasis lesions (80).

### Explanation

All RCTs assess response variables, or outcomes, for which the groups are compared. Most trials have several outcomes, some of which are of more interest than others. The *primary outcome measure* is the prespecified outcome of greatest importance and is usually the one used in the sample size calculation (item 7). Some trials may have more than one primary outcome. Having more than one or two outcomes, however, incurs the problems of interpretation associated with multiplicity\* of analyses (see items 18 and 20) and is not recommended. Primary outcomes should be explicitly indicated as such in the report of an RCT. Other outcomes of interest are *secondary outcomes*. There may be several secondary outcomes, which often include unanticipated or unintended effects of the intervention (item 19).

All outcome measures, whether primary or secondary, should be identified and completely defined. When outcomes are assessed at several time points after randomization, authors should indicate the prespecified time point of primary interest. It is sometimes helpful to specify who assessed outcomes (for example, if special

skills are required to do so) and how many assessors there were.

Many diseases have a plethora of possible outcomes that can be measured by using different scales or instruments. Where available and appropriate, previously developed and validated scales or consensus guidelines should be used (83, 84), both to enhance quality of measurement and to assist in comparison with similar studies. For example, assessment of quality of life is likely to be improved by using a validated instrument (85). Authors should indicate the provenance and properties of scales.

More than 70 outcomes were used in 196 RCTs of nonsteroidal anti-inflammatory drugs for rheumatoid arthritis (28), and 640 different instruments had been used in 2000 trials in schizophrenia, of which 369 had been used only once (33). Investigation of 149 of those 2000 trials showed that unpublished scales were a source of bias. In nonpharmacologic trials, one third of the claims of treatment superiority based on unpublished scales would not have been made if a published scale had been used (86). Similar evidence has been reported elsewhere (87, 88).

*Item 6b.* When applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).

### Examples

The clinical end point committee . . . evaluated all clinical events in a blinded fashion and end points were determined by unanimous decision (89).

Blood pressure (diastolic phase 5) while the patient was sitting and had rested for at least five minutes was measured by a trained nurse with a Copal UA-251 or a Takeda UA-751 electronic auscultatory blood pressure reading machine (Andrew Stephens, Brighouse, West Yorkshire) or with a Hawksley random zero sphygmomanometer (Hawksley, Lancing, Sussex) in patients with atrial fibrillation. The first reading was discarded and the mean of the next three consecutive readings with a coefficient of variation below 15% was used in the study, with additional readings if required (90).

### Explanation

Authors should give full details of how the primary and secondary outcomes were measured and whether any particular steps were taken to increase the reliability of the measurements.

Some outcomes are easier to measure than others. Death (from any cause) is usually easy to assess, whereas blood pressure, depression, or quality of life are more difficult. Some strategies can be used to improve the quality of measurements. For example, assessment of blood pressure is more reliable if more than one reading is obtained, and digit preference can be avoided by using a random-zero sphygmomanometer. Assessments are more likely to be free of bias if the participant and assessor are blinded to group assignment (item 11a). If a trial requires taking unfamiliar measurements, formal, standardized training of the people who will be taking the measurements can be beneficial.

*Item 7a.* How sample size was determined.

#### Examples

We believed that . . . the incidence of symptomatic deep venous thrombosis or pulmonary embolism or death would be 4% in the placebo group and 1.5% in the ardeparin sodium group. Based on 0.9 power to detect a significant difference ( $P = 0.05$ , two-sided), 976 patients were required for each study group. To compensate for nonevaluable patients, we planned to enroll 1000 patients per group (91).

To have an 85% chance of detecting as significant (at the two sided 5% level) a five point difference between the two groups in the mean SF-36 [Short Form-36] general health perception scores, with an assumed standard deviation of 20 and a loss to follow up of 20%, 360 women (720 in total) in each group were required (92).

#### Explanation

For scientific and ethical reasons, the sample size for a trial needs to be planned carefully, with a balance between clinical and statistical considerations. Ideally, a study should be large enough to have a high probability (power) of detecting as statistically significant a clinically important difference of a given size if such a difference exists. The size of effect deemed important is inversely related to the sample size necessary to detect it; that is, large samples are necessary to detect small differences. Elements of the sample size calculation are 1) the estimated outcomes in each group (which implies the clinically important target difference between the intervention groups); 2) the  $\alpha$  (type I) error level; 3) the statistical power (or the  $\beta$  [type II] error level); and 4)

for continuous outcomes, the standard deviation of the measurements (93).

Authors should indicate how the sample size was determined. If a formal power calculation was used, the authors should identify the primary outcome on which the calculation was based (item 6a), all the quantities used in the calculation, and the resulting target sample size per comparison group. It is preferable to quote the postulated results of each group rather than the expected difference between the groups. Details should be given of any allowance made for attrition during the study.

In some trials, interim analyses are used to help decide whether to continue recruiting (item 7b). If the actual sample size differed from that originally intended for some other reason (for example, because of poor recruitment or revision of the target sample size), the explanation should be given.

Reports of studies with small samples frequently include the erroneous conclusion that the intervention groups do not differ, when too few patients were studied to make such a claim (94). Reviews of published trials have consistently found that a high proportion of trials have very low power to detect clinically meaningful treatment effects (17, 95). In reality, small but clinically valuable true differences are likely, which require large trials to detect (96). The median sample size was 54 patients in 196 trials in arthritis (28), 46 patients in 73 trials in dermatology (8), and 65 patients in 2000 trials in schizophrenia (39). Many reviews have found that few authors report how they determined the sample size (8, 14, 25, 39).

There is little merit in calculating the statistical power once the results of the trial are known; the power is then appropriately indicated by confidence intervals\* (item 17) (97).

*Item 7b.* When applicable, explanation of any interim analyses and stopping rules.

#### Examples

The results of the study . . . were reviewed every six months to enable the study to be stopped early if, as indeed occurred, a clear result emerged (98).

Two interim analyses were performed during the trial. The levels of significance maintained an overall  $P$  value of 0.05 and were calculated according to the O'Brien–Fleming stopping boundaries. This final analy-

sis used a Z score of 1.985 with an associated *P* value of 0.0471 (99).

#### Explanation

Many trials recruit participants over a long period. If an intervention is working particularly well or badly, the study may need to be ended early for ethical reasons. This concern can be addressed by examining results as the data accumulate. However, performing multiple statistical examinations of accumulating data without appropriate correction can lead to erroneous results and interpretations (100). If the accumulating data from a trial are examined at five interim analyses\*, the overall false-positive rate is nearer to 19% than to the nominal 5%.

Several group sequential statistical methods are available to adjust for multiple analyses (101–103); their use should be prespecified in the trial protocol. With these methods, data are compared at each interim analysis, and a very small *P* value indicates statistical significance. Some trialists use these *P* values as an aid to decision making (104), whereas others treat them as a formal stopping rule\* (with the intention that the trial will cease if the observed *P* value is smaller than the critical value).

Authors should report whether they took multiple “looks” at the data and, if so, how many there were, the statistical methods used (including any formal stopping rule), and whether they were planned before the initiation of the trial or some time thereafter. This information is frequently not included in published trial reports (14).

*Item 8a.* Method used to generate the random allocation sequence.

#### Example

Independent pharmacists dispensed either active or placebo inhalers according to a computer generated randomization list (62).

#### Explanation

Ideally, participants should be assigned to comparison groups in the trial on the basis of a chance (random) process characterized by unpredictability (Table 1). Authors should provide sufficient information that the reader can assess the methods used to generate the random allocation sequence\* and the likelihood of bias in group assignment.

Many methods of sequence generation are adequate. However, readers cannot judge adequacy from such terms as “random allocation,” “randomization,” or “random” without further elaboration. Authors should specify the method of sequence generation, such as a random-number table or a computerized random-number generator. The sequence may be generated by the process of minimization,\* a method of restricted randomization\* (item 8b) (Table 3).

In some trials, participants are intentionally allocated in unequal numbers to each intervention: for example, to gain more experience with a new procedure or to limit costs of the trial. In such cases, authors should report the randomization ratio (for example, 2:1).

The term *random* has a precise technical meaning. With random allocation, each participant has a known probability of receiving each treatment before one is assigned, but the actual treatment is determined by a chance process and cannot be predicted. However, “random” is often used inappropriately in the literature to describe trials in which nonrandom, “deterministic\*” allocation methods, such as alternation, hospital numbers, or date of birth, were used. When investigators use such a method, they should describe it exactly and should not use the term “random” or any variation of it. Even the term “quasi-random” is questionable for such trials. Empirical evidence (2–5) indicates that such trials give biased results. Bias presumably arises from the inability to conceal these allocation systems adequately (see item 9).

Only 32% of reports published in specialty journals (21) and 48% of reports published in general medical journals (25) specified an adequate method for generating random numbers. In almost all of these cases, researchers used a random-number generator on a computer or a random-number table. A review of one dermatology journal over 22 years found that adequate generation was reported in only 1 of 68 trials (8).

*Item 8b.* Details of any restriction [of randomization] (e.g., blocking, stratification).

#### Example

Women had an equal probability of assignment to the groups. The randomization code was developed using a computer random number generator to select random permuted blocks. The block lengths were 4, 8, and 10 varied randomly . . . (74)

**Table 3. Item 8b: Restricted Randomization**

Randomization based on a single sequence of random assignments (as described in item 8a) is known as *simple randomization*. *Restricted randomization* describes any procedure to control the randomization to achieve balance between groups in size or characteristics. *Blocking* is used to ensure that comparison groups will be of approximately the same size; *stratification* is used to ensure good balance of participant characteristics in each group.

**Blocking**

Blocking can be used to ensure close balance of the numbers in each group at any time during the trial. After a block of every 10 participants was assigned, for example, 5 would be allocated to each arm of the trial (105). Improved balance comes at the cost of reducing the unpredictability of the sequence. Although the order of interventions varies randomly within each block, a person running the trial could deduce some of the next treatment allocations if they discovered the block size (106). Blinding the interventions, using larger block sizes, and randomly varying the block size can ameliorate this problem.

**Stratification**

By chance, particularly in small trials, study groups may not be well matched for baseline characteristics\*, such as age and stage of disease. This weakens the trial's credibility (107). Such imbalances can be avoided without sacrificing the advantages of randomization. Stratification ensures that the numbers of participants receiving each intervention are closely balanced within each stratum. *Stratified randomization\** is achieved by performing a separate randomization procedure within each of two or more subsets of participants (for example, those defining age, smoking, or disease severity). Stratification by center is common in multicenter trials. Stratification requires blocking within strata; without blocking, it is ineffective.

**Minimization**

Minimization ensures balance between intervention groups for several patient factors (32, 59). Randomization lists are not set up in advance. The first patient is truly randomly allocated; for each subsequent patient, the treatment allocation is identified, which minimizes the imbalance between groups at that time. That allocation may then be used, or a choice may be made at random with a heavy weighting in favor of the intervention that would minimize imbalance (for example, with a probability of 0.8). The use of a random component is generally preferable. Minimization has the advantage of making small groups closely similar in terms of participant characteristics at all stages of the trial.

Minimization offers the only acceptable alternative to randomization, and some have argued that it is superior (108). Trials that use minimization are considered methodologically equivalent to randomized trials, even when a random element is not incorporated.

Terms marked with an asterisk are defined in the glossary at the end of the text.

**Explanation**

In large trials, simple randomization\* can be trusted to generate similar numbers in the two trial groups and to generate groups that are roughly comparable in terms of known (and unknown) prognostic variables\*. *Restricted randomization\** describes procedures used to control the randomization to achieve balance between groups in size or characteristics (Table 3).

It is helpful to indicate whether no restriction was used, such as by stating that “simple randomization” was done. Otherwise, the methods used to restrict the randomization, along with the method used for random selection (item 8a), should be specified. For block ran-

domization, authors should provide details on how the blocks were generated (for example, by using a permuted block design\*), the block size or sizes, and whether the block size was randomly varied. Authors should specify whether stratification was used, and if so, which factors were involved and the methods used for blocking. Although stratification is a useful technique, especially for smaller trials, it is complicated to implement if many stratifying factors are used. If minimization (Table 3) was used, it should be explicitly identified, as should the variables incorporated into the scheme. Use of a random element should be indicated.

Stratification has been shown to increase the power of small randomized trials by up to 12%, especially in the presence of a large intervention effect or strong prognostic stratifying variables (109). Minimization does not provide the same advantage (110).

Only 9% of 206 reports of trials in specialty journals (21) and 39% of 80 trials in general medical journals reported use of stratification (25). In each case, only about half of the reports mentioned the use of restricted randomization. Those studies and that of Adetugbo and Williams (8) found that the sizes of the treatment groups in many trials were very often the same or quite similar, yet blocking or stratification had not been mentioned. One possible cause of this close balance in numbers is underreporting of the use of restricted randomization.

*Item 9.* Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.

**Example**

Women were assigned on an individual basis to both vitamins C and E or to both placebo treatments. They remained on the same allocation throughout the pregnancy if they continued in the study. A computer-generated randomisation list was drawn up by the statistician . . . and given to the pharmacy departments. The researchers responsible for seeing the pregnant women allocated the next available number on entry into the trial (in the ultrasound department or antenatal clinic), and each woman collected her tablets direct from the pharmacy department. The code was revealed to the researchers once recruitment, data collection, and laboratory analyses were complete (111).

*Explanation*

Item 8 discussed generation of an unpredictable sequence of assignments. Of considerable importance is how this sequence is applied when participants are enrolled into the trial. A generated allocation schedule should ideally be implemented by using allocation concealment (21), a critical process that prevents foreknowledge of treatment assignment and thus shields those who enroll participants from being influenced by this knowledge. The decision to accept or reject a participant should be made, and informed consent should be obtained from the participant, in ignorance of the next assignment in the sequence (112).

Allocation concealment should not be confused with blinding (item 11). Allocation concealment seeks to prevent selection bias, protects the assignment sequence before and until allocation, and can always be successfully implemented (2). In contrast, blinding seeks to prevent performance\* and ascertainment bias\*, protects the sequence after allocation, and cannot always be implemented (21). Without adequate allocation concealment, however, even random, unpredictable assignment sequences can be subverted (2, 113).

Decentralized or “third-party” assignment is especially desirable. Many good approaches to allocation concealment incorporate external involvement. Use of a pharmacy or central telephone randomization system are two common techniques. Automated assignment systems are likely to become more common (114). When external involvement is not feasible, an excellent method of allocation concealment is the use of numbered containers. The interventions (often medicines) are sealed in sequentially numbered identical containers according to the allocation sequence. Enclosing assignments in sequentially numbered, opaque, sealed envelopes can be a good allocation concealment mechanism if it is developed and monitored diligently. This method can be corrupted, particularly if it is poorly executed. Investigators should ensure that the envelopes are opened sequentially and only after the participant’s name and other details are written on the appropriate envelope (106).

Recent studies provide empirical evidence of bias leaking into trials. Investigators assessed the quality of reporting of randomization in 250 controlled trials extracted from 33 meta-analyses of topics in pregnancy and childbirth, and then analyzed the associations between those assessments and the estimated effects of the

intervention (2). Trials in which the allocation sequence had been inadequately or unclearly concealed yielded larger estimates of treatment effects (odds ratios were exaggerated, on average, by 30% to 40%) than did trials in which authors reported adequate allocation concealment. Three other studies (3–5) had similar results. These findings provide strong empirical evidence that inadequate allocation concealment contributes to bias in estimating treatment effects.

Despite the importance of the mechanism of allocation, published reports frequently omit such details. The mechanism used to allocate interventions was omitted in reports of 89% of trials in rheumatoid arthritis (28), 48% of trials in obstetrics and gynecology journals (21), and 44% of trials in general medical journals (25). Only 5 of 73 reports of RCTs published in one dermatology journal between 1976 and 1997 reported the method used to allocate treatments (8).

*Item 10.* Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.

*Example*

Determination of whether a patient would be treated by streptomycin and bed-rest (S case) or by bed-rest alone (C case) was made by reference to a statistical series based on random sampling numbers drawn up for each sex at each centre by Professor Bradford Hill; the details of the series were unknown to any of the investigators or to the co-ordinator and were contained in a set of sealed envelopes, each bearing on the outside only the name of the hospital and a number. After acceptance of a patient by the panel, and before admission to the streptomycin centre, the appropriate numbered envelope was opened at the central office; the card inside told if the patient was to be an S or a C case, and this information was then given to the medical officer of the centre (33).

*Explanation*

As noted in item 9, concealment of the allocated intervention at the time of enrollment is especially important. Thus, in addition to knowing the methods used, it is also important to understand how the random sequence was implemented: specifically, who generated the allocation sequence, who enrolled participants, and who assigned participants to trial groups.

The process of enrolling participants into a trial has

**Table 4. Generation and Implementation of a Random Sequence of Treatments**

Generation	Implementation
Preparation of the random sequence	Enrolling participants Assessing eligibility Discussing the trial Obtaining informed consent Enrolling patient in trial
Preparation of an allocation system (such as coded bottles or envelopes), preferably designed to be concealed from the person assigning participants to groups	Ascertaining treatment assignment (such as by opening the next envelope) Administering intervention

two very different aspects: generation and implementation (Table 4). Although the same persons may carry out more than one process under each heading, investigators should strive for complete separation of the people involved in the generation and implementation of assignments.

Whatever the methodologic quality of the randomization process, failure to separate creation of the allocation sequence from assignment to study group may introduce bias. For example, the person who generated an allocation sequence could retain a copy and consult it when interviewing potential participants for a trial. Thus, that person could bias the enrollment\* or assignment process, regardless of the unpredictability of the assignment sequence. Nevertheless, the same person may sometimes have to prepare the scheme and also be involved in group assignment. Investigators must then ensure that the assignment schedule is unpredictable and locked away from even the person who generated it. The report of the trial should specify where the investigators stored the allocation list.

*Item 11a.* Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment.

#### Example

All study personnel and participants were blinded to treatment assignment for the duration of the study. Only the study statisticians and the data monitoring committee saw unblinded data, but none had any contact with study participants (115).

#### Explanation

In controlled trials, the term *blinding\** refers to keeping study participants, health care providers, and sometimes those collecting and analyzing clinical data unaware of the assigned intervention, so that they will not be influenced by that knowledge. Blinding is important to prevent bias at several stages of a controlled trial, although its relevance varies according to circumstances.

Blinding of patients is important because knowledge of group assignment may influence responses to treatment. Patients who know that they have been assigned to receive the new treatment may have favorable expectations or increased anxiety. Patients assigned to standard treatment may feel discriminated against or reassured. Use of placebo controls coupled with blinding of patients is intended to prevent bias resulting from nonspecific effects associated with receiving the intervention (placebo effects).

Blinding of patients and health care providers prevents performance bias. This type of bias can occur if additional therapeutic interventions (sometimes called “co-interventions”) are provided or sought preferentially by trial participants in one of the comparison groups. The decision to withdraw a participant from a study or to adjust the dose of medication could easily be influenced by knowledge of the participant’s group assignment.

Blinding of patients, health care providers, and other persons (for example, radiologists) involved in evaluating outcomes minimizes the risk for detection bias, also called observer, ascertainment, or assessment bias. This type of bias occurs if knowledge of a patient’s assignment influences the process of outcome assessment. For example, in a placebo-controlled multiple sclerosis trial, assessments by unblinded, but not blinded, neurologists showed an apparent benefit of the intervention (116).

Finally, blinding of the data analyst can also prevent bias. Knowledge of the interventions received may influence the choice of analytical strategies and methods (117).

Trials without any blinding are known as “open\*” or, if they are pharmaceutical trials, “open-label.” This design is common in early investigations of a drug (phase II trials).

Unlike allocation concealment (item 10), blinding may not always be appropriate or possible. An example

is a trial comparing levels of pain associated with sampling blood from the ear or thumb (118). Blinding is particularly important when outcome measures involve some subjectivity, such as assessment of pain or cause of death. It is less important for objective criteria, such as death from any cause, when there is little scope for ascertainment bias. Even then, however, lack of blinding in any trial can lead to other problems, such as attrition (Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding. Submitted for publication). In certain trials, especially surgical trials, double-blinding is difficult or impossible. However, blinded assessment of outcome can often be achieved even in open trials. For example, lesions can be photographed before and after treatment and be assessed by someone not involved in performance of the trial (119). Some treatments have unintended effects that are so specific that their occurrence will inevitably identify the treatment received to both the patient and the medical staff. Blinded assessment of outcome is especially useful when such revelation is a risk.

Many trials are described as “double blind.” Although this term implies that neither the caregiver nor the patient knows which treatment was received, it is ambiguous with regard to blinding of other persons, including those assessing patient outcome (120). Authors should state who was blinded (for example, participants, care providers, evaluators, monitors, or data analysts), the mechanism of blinding (for example, capsules or tablets), and the similarity of characteristics of treatments (for example, appearance, taste, and method of administration) (40, 121). They should also explain why any participants, care providers, or evaluators were not blinded.

Authors frequently do not report whether or not blinding was used (16), and when blinding is specified, details are often missing. For example, reports of 51% of 506 trials in cystic fibrosis (122), 33% of 196 trials in rheumatoid arthritis (28), and 38% of 68 trials in dermatology (8) did not state whether blinding was used. Of 31 “double-blind” trials in obstetrics and gynecology, only 14 (45%) of the reports indicated the similarity of the treatment and control regimens. Moreover, only 5 (16%) stated explicitly that blinding had been successful (121).

The term *masking* is sometimes used in preference to *blinding* to avoid confusion with the medical condi-

tion of being without sight. However, “blinding” in its methodologic sense appears to be understood worldwide and is acceptable for reporting clinical trials (119, 123).

*Item 11b.* If done, how the success of blinding was evaluated.

#### Example

To evaluate patient blinding, the questionnaire asked patients to indicate which treatment they believed they had received (acupuncture, placebo, or don't know) at 3 points in time . . . If patients answered either acupuncture or placebo, they were asked to indicate what led to that belief . . . (124).

#### Explanation

Just as we seek evidence of concealment to assure us that assignment was truly random, we may seek evidence that blinding was successful. Although description of the mechanism used for blinding may provide such assurance, the success of blinding can sometimes be evaluated directly by asking participants, caregivers, or outcome assessors which treatment they think they received.

Prasad and colleagues (63) reported a placebo-controlled trial of zinc lozenges for reducing the duration of symptoms of the common cold. They carried out a separate study in healthy volunteers to check the comparability of taste of zinc or placebo lozenges. They also asked participants in the main trial to try to identify which treatment they were receiving. They reported that at the end of the trial, 56% of the zinc recipients and 26% of the placebo recipients correctly identified their group assignment ( $P = 0.09$ ).

In principle, if blinding was successful, the ability of participants to accurately guess their group assignment should be no better than chance. In practice, however, if participants do successfully identify their assigned intervention more often than expected by chance, it may not mean that blinding was unsuccessful. Although adverse effects in particular may offer strong clues as to which intervention was received, especially in studies of pharmacologic agents, the clinical outcome may also provide clues. Thus, clinicians are likely to assume, not always correctly, that a patient who had a favorable outcome was more likely to have received the active intervention rather than control. If the active intervention is indeed

beneficial, their “guesses” would be likely to be better than those produced by chance (125).

Authors should report any failure of the blinding procedure, such as placebo and active preparations that were not identical in appearance.

*Item 12a.* Statistical methods used to compare groups for primary outcome(s).

#### Example

All data analysis was carried out according to a pre-established analysis plan. Proportions were compared by using  $\chi^2$  tests with continuity correction or Fisher's exact test when appropriate. Multivariate analyses were conducted with logistic regression. The durations of episodes and signs of disease were compared by using proportional hazards regression. Mean serum retinol concentrations were compared by *t* test and analysis of covariance . . . Two sided significance tests were used throughout (126).

#### Explanation

Data can be analyzed in many ways, some of which may not be strictly appropriate in a particular situation. It is essential to specify which statistical procedure was used for each analysis, and further clarification may be necessary in the results section of the report.

Almost all methods of analysis yield an estimate of the treatment effect, which is a contrast between the outcomes in the comparison groups. In addition, authors should present a confidence interval for the estimated effect, which indicates a range of uncertainty for the true treatment effect. The confidence interval may also be interpreted as the range of values for the treatment effect that is compatible with the observed data. It is customary to present a 95% confidence interval, which gives the range of uncertainty expected to include the true value in 95 of 100 similar studies.

Study findings can also be assessed in terms of their statistical significance. The *P* value represents the probability that the observed data (or a more extreme result) could have arisen by chance when the interventions did not differ. Actual *P* values (for example, *P* = 0.003) are preferred to imprecise threshold reports (*P* < 0.05) (46, 127).

Standard methods of analysis assume that the data are “independent.” For controlled trials, this usually means that there is one observation per participant.

Treating multiple observations from one participant as independent data is a serious error; such data are produced when outcomes can be measured on different parts of the body, as in dentistry or rheumatology. Data analysis should be based on counting each participant once (128, 129) or should be done by using more complex statistical procedures (130). Incorrect analysis of multiple observations was seen in 123 (63%) of 196 trials in rheumatoid arthritis (28).

*Item 12b.* Methods for additional analyses, such as subgroup analyses and adjusted analyses.

#### Examples

Proportions of patients responding were compared between treatment groups with the Mantel-Haenszel  $\chi^2$  test, adjusted for the stratification variable, methotrexate use (80).

. . . it was planned to assess the relative benefit of CHART in an exploratory manner in subgroups: age, sex, performance status, stage, site, and histology. To test for differences in the effect of CHART, a chi-squared test for interaction was performed, or when appropriate a chi-squared test for trend (131).

#### Explanation

As is the case for primary analyses, the method of subgroup analysis\* should be clearly specified. The strongest analyses are those based on looking for evidence of a difference in treatment effect in complementary subgroups (for example, older and younger participants), a comparison known as a *test of interaction\** (132, 133). A common but inferior approach is to compare *P* values for separate analyses of the treatment effect in each group. It is incorrect to infer a subgroup effect (interaction) from one significant and one nonsignificant *P* value (134). Such inferences have a high false-positive rate.

Because of the high risk for spurious findings, subgroup analyses are often discouraged (14, 135). Post hoc subgroup comparisons (analyses done after looking at the data) are especially likely not to be confirmed by further studies. Such analyses do not have great credibility.

In some studies, imbalances in participant characteristics (prognostic variables) are adjusted\* for by using some form of multiple regression analysis. Although the need for adjustment is much less in RCTs than in epidemiologic studies, an adjusted analysis may be sensible,

especially if one or more prognostic variables seem important (136). Ideally, adjusted analyses should be specified in the study protocol. For example, adjustment is often recommended for any stratification variables (item 8b). In RCTs, the decision to adjust should not be determined by whether baseline differences are statistically significant (133, 137) (item 16). The rationale for any adjusted analyses and the statistical methods used should be specified.

Authors should clarify the choice of variables that were adjusted for, indicate how continuous variables were handled, and specify whether the analysis was planned\* or suggested by the data (Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross sectional survey. Submitted for publication). Reviews of published studies show that reporting of adjusted analyses is inadequate with regard to all of these aspects (138–140).

## Results

*Item 13a.* Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome.

### Examples

See **Figures 2, 3, and 4.**

### Explanation

The design and execution of some RCTs is straightforward, and the flow of participants through each phase of the study can be described adequately in a few sentences. In more complex studies, it may be difficult for readers to discern whether and why some participants did not receive the treatment as allocated, were lost to follow-up\*, or were excluded from the analysis (54). This information is crucial for several reasons. Participants who were excluded after allocation are unlikely to be representative of all participants in the study. For example, patients may not be available for follow-up evaluation because they experienced an acute exacerbation of their illness or severe side effects\* of treatment (32, 141).

Attrition as a result of loss to follow up, which is often unavoidable, needs to be distinguished from inves-

tigator-determined exclusion for such reasons as ineligibility, withdrawal from treatment, and poor adherence to the trial protocol. Erroneous conclusions can be reached if participants are excluded from analysis, and imbalances in such omissions between groups may be especially indicative of bias (141–143). Information about whether the investigators included in the analysis all participants who underwent randomization, in the groups to which they were originally allocated (intention-to-treat analysis [item 16]), is therefore of particular importance. Knowing the number of participants who did not receive the intervention as allocated or did not complete treatment permits the reader to assess to what extent the estimated efficacy of therapy might be underestimated in comparison with ideal circumstances. If available, the number of persons assessed for eligibility should also be reported. Although this number is relevant to external validity only and is arguably less important than the other counts (55), it is a useful indicator of whether trial participants were likely to be representative of all eligible participants.

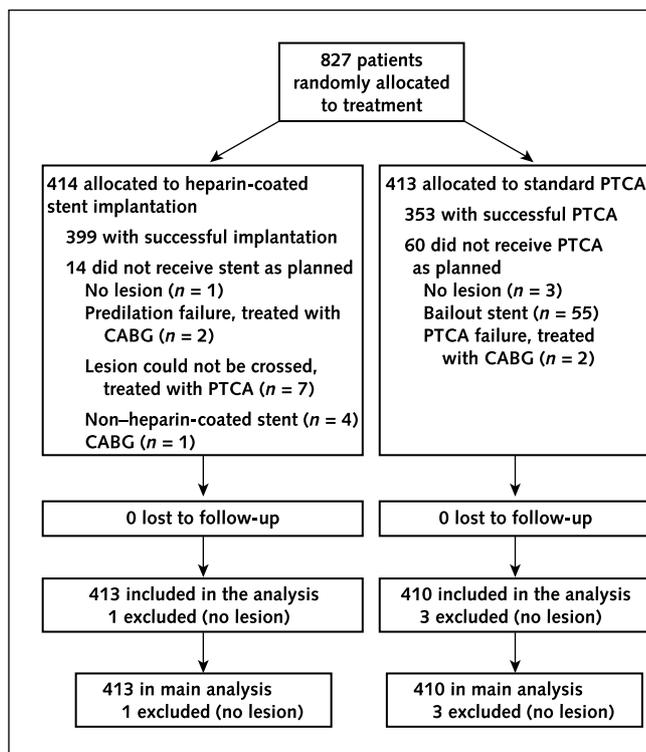
A recent review of RCTs published in five leading general and internal medicine journals in 1998 found that reporting of the flow of participants was often incomplete, particularly with regard to the number of participants receiving the allocated intervention and the number lost to follow-up (54). Even information as basic as the number of participants who underwent randomization and the number excluded from analyses was not available in up to 20% of articles (54). Reporting was considerably more thorough in articles that included a diagram of the flow of participants through a trial, as recommended by CONSORT. This study informed the design of the revised flow diagram in the revised CONSORT statement (56–58). The suggested template is shown in **Figure 1**, and the counts required are described in detail in **Table 5**.

Some information, such as the number of persons assessed for eligibility, may not always be known (14), and depending on the nature of a trial, some counts may be more relevant than others. It will therefore often be useful or necessary to adapt the structure of the flow diagram to a particular trial. For example, a multicenter trial compared implantation of heparin-coated stents with standard percutaneous transluminal angioplasty in patients scheduled to undergo coronary angioplasty (144). The nature of the intervention meant that a rel-

**Table 5. Information Required To Document the Flow of Participants through Each Stage of a Randomized, Controlled Trial**

Stage	Number of People Included	Number of People Not Included or Excluded	Rationale
Enrollment	People evaluated for potential enrollment	People who did not meet the inclusion criteria People who met the inclusion criteria but declined to be enrolled	These counts indicate whether trial participants were likely to be representative of all patients seen; they are relevant to assessment of external validity only, and they are often not available
Randomization	Participants randomly assigned		Crucial count for defining trial size and assessing whether a trial has been analyzed by intention to treat
Treatment allocation	Participants who received treatment as allocated, by study group	Participants who did not receive treatment as allocated, by study group	Important counts for assessment of internal validity and interpretation of results; reasons for not receiving treatment as allocated should be given
Follow-up	Participants who completed treatment as allocated, by study group Participants who completed follow-up as planned, by study group	Participants who did not complete treatment as allocated, by study group Participants who did not complete follow-up as planned, by study group	Important counts for assessment of internal validity and interpretation of results; reasons for not completing treatment or follow-up should be given
Analysis	Participants included in main analysis, by study group	Participants excluded from main analysis, by study group	Crucial count for assessing whether a trial has been analyzed by intention to treat; reasons for excluding participants should be given

**Figure 2. Flow diagram of a multicenter trial comparing implantation of heparin-coated stents with percutaneous transluminal angioplasty (PTCA).**



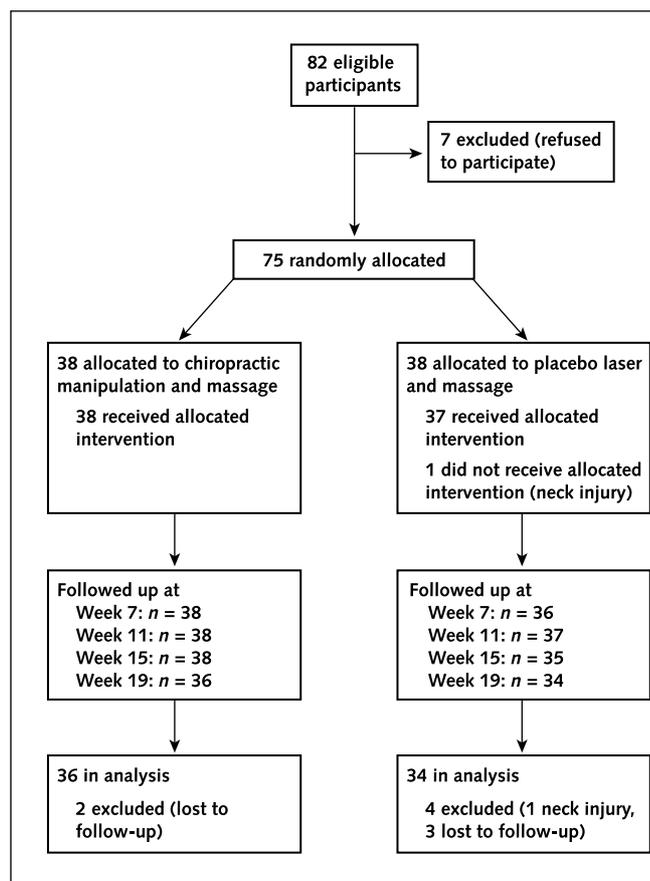
The diagram includes detailed information on the interventions received. CABG = coronary artery bypass grafting. Adapted from reference 144.

atively large number of patients did not receive the allocated intervention. In the flow diagram (Figure 2), the box describing treatment allocation had to be expanded to reflect this.

In some situations, other information may usefully be added. For example, the flow diagram of a trial of chiropractic manipulation of the cervical spine in the treatment of episodic tension-type headache (145) showed the number of patients actively followed up at different times during the study (Figure 3). The main results, such as the number of events for the primary outcome, may sometimes be added to the flow diagram. For example, the flow diagram of a trial of the topoisomerase I inhibitor irinotecan in patients with metastatic colorectal cancer in whom fluorouracil chemotherapy had failed (146) included the number of deaths (Figure 4).

These examples illustrate that the exact form and content of the flow diagram may be varied according to specific features of a trial. For example, many trials of surgery or vaccination do not include the possibility of discontinuation. Although CONSORT strongly recommends using this graphical device to communicate participant flow throughout the study, there is no specific, prescribed format. Inclusion of a diagram may be unnecessary for simple trials without losses to follow-up or exclusions.

**Figure 3.** Flow diagram of a trial of chiropractic manipulation of the cervical spine for treatment of episodic tension-type headache.



The diagram includes the number of patients actively followed up at different times during the trial. Adapted from reference 145.

*Item 13b.* Describe protocol deviations from study as planned, together with reasons.

**Examples**

There was only one protocol deviation, in a woman in the study group. She had an abnormal pelvic measurement and was scheduled for elective caesarean section. However, the attending obstetrician judged a trial of labour acceptable; caesarean section was done when there was no progress in the first stage of labour (147).

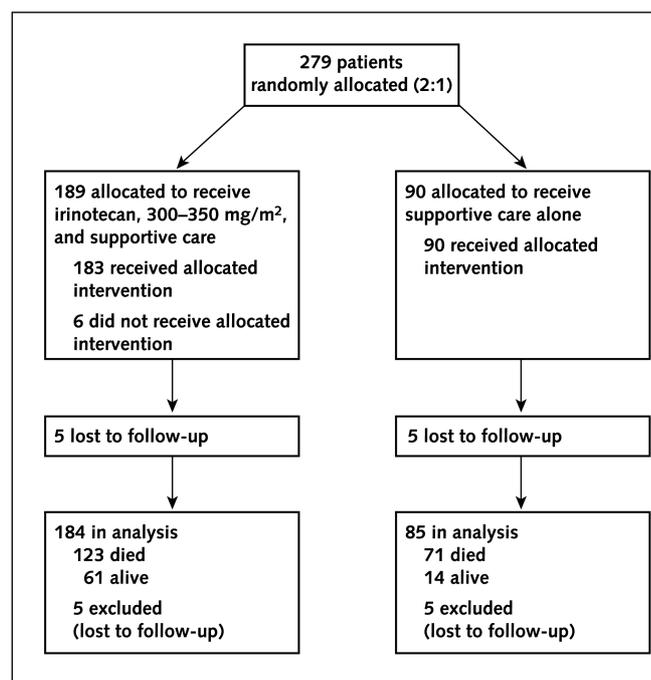
The monitoring led to withdrawal of nine centres, in which existence of some patients could not be proved, or other serious violations of good clinical practice had occurred (148).

**Explanation**

Authors should report all departures from the protocol, including unplanned changes to interventions, examinations, data collection, and methods of analysis. Some of these protocol deviations\* may be reported in the flow diagram (item 13a): for example, participants who did not receive the intended intervention. If participants were excluded after randomization because they were found not to meet eligibility criteria (item 16) (contrary to the intention-to-treat principle), they can be included in the flow diagram. Use of the term “protocol deviation” in published articles is not sufficient to justify exclusion of participants after randomization. The nature of the protocol deviation and the exact reason for excluding participants after randomization should always be reported.

*Item 14.* Dates defining the periods of recruitment and follow-up.

**Figure 4.** Flow diagram of a trial of the topoisomerase I inhibitor irinotecan in patients with metastatic colorectal cancer in whom fluorouracil chemotherapy had failed.



The diagram includes the results for the main outcome (overall survival). Adapted from reference 146.

**Example**

Age-eligible participants were recruited . . . from February 1993 to September 1994 . . . Participants attended clinic visits at the time of randomization (baseline) and at 6-month intervals for 3 years (115).

**Explanation**

Knowing when a study took place and over what period participants were recruited places the study in historical context. Medical and surgical therapies, including concurrent therapies, evolve continuously and may affect the routine care given to patients during a trial. Knowing the rate at which participants were recruited may also be useful, especially to other investigators.

The length of follow-up is not always a fixed period after randomization. In many RCTs in which the outcome is time to an event, follow-up of all participants is ended on a specific date. This date should be given, and it is also useful to report the median duration of follow-up (149, 150).

If the trial was stopped owing to results of interim analysis of the data (item 7b), this should be reported. Early stopping will lead to a discrepancy between the planned and actual sample sizes. In addition, trials that stop early are likely to overestimate the treatment effect (102).

In a review of reports in oncology journals that used survival analysis, most of which were not RCTs, Altman and associates (150) found that nearly 80% (104 of 132

reports) included the starting and ending dates for accrual of patients, but only 24% (32 of 132 reports) also reported the date on which follow-up ended.

*Item 15.* Baseline demographic and clinical characteristics of each group.

**Example**

See **Table 6**.

**Explanation**

Although the eligibility criteria (item 3) indicate who was eligible for the trial, it is also important to know the characteristics of the participants who were actually recruited. This information allows readers, especially clinicians, to judge how relevant the results of a trial might be to a particular patient.

Randomized, controlled trials aim to compare groups of participants that differ only with respect to the intervention (treatment). Although proper random assignment prevents selection bias, it does not guarantee that the groups are equivalent at baseline. Any differences in baseline characteristics are, however, the result of chance rather than bias (25). The study groups should be compared at baseline for important demographic and clinical characteristics so that readers can assess how comparable the groups were. Baseline data may be especially valuable when the outcome measure can also be measured at the start of the trial.

Baseline information is efficiently presented in a table (**Table 6**). For continuous variables, such as weight or blood pressure, the variability of the data should be reported, along with average values. Continuous variables can be summarized for each group by the mean and standard deviation. When continuous data have an asymmetrical distribution, a preferable approach may be to quote the median and a percentile range (perhaps the 25th and 75th percentiles) (127). Standard errors and confidence intervals are not appropriate for describing variability—they are inferential rather than descriptive statistics. Variables making up a small number of ordered categories (such as stages of disease I to IV) should not be treated as continuous variables; instead, numbers and proportions should be reported for each category (46, 127).

Despite many warnings about their inappropriateness (21, 25, 151) significance tests of baseline differences are still common; they were reported in half of the

**Table 6. Item 15: Example of Reporting of Baseline Demographic and Clinical Characteristics of Trial Group†**

Characteristic	Vitamin Group (n = 141)	Placebo Group (n = 142)
Mean age ± SD, y	28.9 ± 6.4	29.8 ± 5.6
Smokers, n (%)	22 (15.6)	14 (9.9)
Mean body mass index ± SD, kg/m <sup>2</sup>	25.3 ± 6.0	25.6 ± 5.6
Mean blood pressure ± SD, mm Hg		
Systolic	112 ± 15	110 ± 12
Diastolic	67 ± 11	68 ± 10
Parity, n (%)		
0	91 (65)	87 (61)
1	39 (28)	42 (30)
2	9 (6)	8 (6)
>2	2 (1)	5 (4)
Coexisting disease, n (%)		
Essential hypertension	10 (7)	7 (5)
Lupus or antiphospholipid syndrome	4 (3)	1 (1)
Diabetes	2 (1)	3 (2)

† Adapted from part of Table 1 of reference 111.

trials in a recent survey of 50 RCTs (133). Ideally, the trial protocol should state whether or not adjustment is made for nominated baseline variables by using analysis of covariance (137). Adjustment for variables because they differ significantly at baseline is likely to bias the estimated treatment effect (137).

*Item 16.* Number of participants (denominator) in each group included in each analysis and whether the analysis was by “intention to treat.” State the results in absolute numbers when feasible (e.g., 10 of 20, not 50%).

#### Examples

The primary analysis was intention-to-treat and involved all patients who were randomly assigned . . . (91).

One patient in the alendronate group was lost to follow up; thus data from 31 patients were available for the intention-to-treat analysis. Five patients were considered protocol violators . . . consequently 26 patients remained for the per-protocol analyses (152).

#### Explanation

The number of participants in each group is an essential element of the results. Although the flow diagram may indicate the numbers of participants for whom outcomes were available, these numbers may vary for different outcome measures. The sample size per group (the denominator when reporting proportions) should be given for all summary information. This information is especially important for binary outcomes, because effect measures (such as risk ratio and risk difference) should be interpreted in relation to the event rate. Expressing results as fractions also aids the reader in assessing whether all randomly assigned participants were included in an analysis, and if not, how many were excluded. It follows that results should not be presented solely as summary measures, such as relative risks.

Failure to include all participants may bias trial results. Most trials do not yield perfect data, however. “Protocol violations” may occur, such as when patients do not receive the full intervention or the correct intervention or a few ineligible patients are randomly allocated in error. One widely recommended way to handle such issues is to analyze all participants according to their original group assignment, regardless of what subsequently occurred. This “intention-to-treat” strategy is

not always straightforward to implement. It is common for some patients not to complete a study—they may drop out or be withdrawn from active treatment—and thus are not assessed at the end. Although those participants cannot be included in the analysis, it is customary still to refer to analysis of all available participants as an intention-to-treat analysis. The term is often inappropriately used when some participants for whom data are available are excluded: for example, those who received none of the intended treatment because of nonadherence to the protocol. Conversely, analysis can be restricted to only participants who fulfill the protocol in terms of eligibility, interventions, and outcome assessment. This analysis is known as an “on-treatment” or “per protocol” analysis. Sometimes both types of analysis are presented.

Excluding participants from the analysis can lead to erroneous conclusions. For example, in a trial that compared medical with surgical therapy for carotid stenosis, analysis limited to participants who were available for follow-up showed that surgery reduced the risk for transient ischemic attack, stroke, and death. However, intention-to-treat analysis based on all participants as originally assigned did not show a superior effect of surgery (153). Intention-to-treat analysis is generally favored because it avoids bias associated with nonrandom loss of participants (154–156). Regardless of whether authors use the term “intention to treat,” they should make clear which participants are included in each analysis (item 13). Intention-to-treat analysis is not appropriate for examining adverse effects.

Noncompliance with assigned therapy may mean that the intention-to-treat analysis underestimates the real benefit of the treatment; additional analyses may therefore be considered (157, 158).

In a review of RCTs published in leading general medical journals in 1997, about half of the reports (119 of 249) mentioned intention-to-treat analysis, but only five stated explicitly that all participants who underwent random allocation were analyzed according to group assignment (18). Moreover, 89 (75%) of these trials were missing some data on the primary outcome variable. Schulz and associates (121) found that trials with no reported exclusions were methodologically weaker in other respects than those that reported on some excluded participants, strongly indicating that at least some researchers who had excluded participants did not

**Table 7. Item 17: Example of Reporting of Summary Results for Each Study Group†**

End Point	Etanercept Group (n = 30)	Placebo Group (n = 30)	Difference (95% CI)	P Value
	←————— n (%) —————→		%	
Primary				
Achieved psoriatic arthritis response criteria at 12 weeks	26 (87)	7 (23)	63 (44–83)	<0.001
Secondary				
Proportion of patients meeting ACR criteria				
ACR20	22 (73)	4 (13)	60 (40–80)	<0.001
ACR50	15 (50)	1 (3)	47 (28–66)	<0.001
ACR70	4 (13)	0 (0)	13 (1–26)	0.04

† See also example for item 6a. Adapted from Table 2 of reference 80. ACR = American College of Rheumatology.

report it. Ruiz-Canela and colleagues (159) found that reporting an intention-to-treat analysis was associated with some other aspects of good study design and reporting, such as describing a sample size calculation.

*Item 17.* For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g., 95% confidence interval).

#### Example

See **Table 7**.

#### Explanation

For each outcome, study results should be reported as a summary of the outcome in each group (for example, the proportion of participants with or without the event, or the mean and standard deviation of measurements), together with the contrast between the groups, known as the *effect size*\*. For binary outcomes, the measure of effect could be the risk ratio (relative risk), odds ratio, or risk difference; for survival time data, the measure could be the hazard ratio or difference in median survival time; and for continuous data, it is usually the difference in means. Confidence intervals should be presented for the contrast between groups. A common error is the presentation of separate confidence intervals for the outcome in each group rather than for the treatment effect (160). Trial results are often more clearly displayed in a table rather than in the text, as shown in **Table 7**.

For all outcome measures, authors should provide a confidence interval to indicate the precision\* (uncertainty) of the estimate (46, 161). A 95% confidence interval is conventional, but occasionally other levels are used. Many journals require or strongly encourage the use of confidence intervals (162). They are especially valuable

in relation to nonsignificant differences, for which they often indicate that the result does not rule out an important clinical difference. The use of confidence intervals has increased markedly in recent years, although not in all medical specialties (160). Although *P* values may be provided in addition to confidence intervals, results should not be reported solely as *P* values (163, 164).

Results should be reported for all planned primary and secondary end points, not just for analyses that were statistically significant. As yet, there is little empirical evidence of within-study selective reporting (28), but it is probably a widespread and serious problem (165, 166). In trials in which interim analyses were performed, interpretation should focus on the final results at the close of the trial, not the interim results (167).

For both binary and survival time data, expressing the results also as the number needed to treat for benefit (NNT<sub>B</sub>) or harm (NNT<sub>H</sub>) can be helpful (item 21) (168, 169).

*Item 18.* Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.

#### Example

Another interesting finding was the evidence of some interaction between treatment with vitamin A and severity of disease on presentation, with results slightly in favour of the vitamin A group among patients initially admitted to hospital, the opposite occurring among those treated as outpatients. Although this finding comes from a subgroup analysis which was preplanned, in no case did the different response between the treatment groups reach significance at the 5% level (126).

### Explanation

Multiple analyses of the same data create a considerable risk for false-positive findings (170). Authors should especially resist the temptation to perform many subgroup analyses (133, 135, 171). Analyses that were prespecified in the trial protocol are much more reliable than those suggested by the data. Authors should indicate which analyses were prespecified. If subgroup analyses were undertaken, authors should report which subgroups were examined and why, although presentation of detailed results may not be necessary in all cases. Selective reporting of subgroup analyses could lead to bias (172). Formal evaluations of interaction (item 12b) should be reported as the estimated difference in the intervention effect in each subgroup (with a confidence interval), not just as *P* values.

Assmann and colleagues (133) found that 35 of 50 trial reports included subgroup analyses, of which only 42% used tests of interaction. They noted that it was often difficult to determine whether subgroup analyses had been specified in the protocol.

Similar recommendations apply to analyses in which adjustment was made for baseline variables. If done, both unadjusted and adjusted analyses should be reported. Authors should indicate whether adjusted analyses, including the choice of variables to adjust for, were planned.

*Item 19.* All important adverse events or side effects in each intervention group.

### Example

The proportion of patients experiencing any adverse event was similar between the rBPI21 [recombinant bactericidal/permeability-increasing protein] and placebo groups: 168 (88.4%) of 190 and 180 (88.7%) of 203, respectively, and it was lower in patients treated with rBPI21 than in those treated with placebo for 11 of 12 body systems. . . . the proportion of patients experiencing a severe adverse event, as judged by the investigators, was numerically lower in the rBPI21 group than the placebo group: 53 (27.9%) of 190 versus 74 (36.5%) of 203 patients, respectively. There were only three serious adverse events reported as drug-related and they all occurred in the placebo group (173).

### Explanation

Most interventions have unintended and often undesirable effects in addition to intended effects. Readers

need information about the harms as well as the benefits of interventions to make rational and balanced decisions. The existence and nature of adverse effects can have a major impact on whether a particular intervention will be deemed acceptable and useful. Not all reported adverse events\* observed during a trial are necessarily a consequence of the intervention; some may be a consequence of the condition being treated. Randomized, controlled trials offer the best approach for providing safety data as well as efficacy data, although they cannot detect rare adverse effects.

At a minimum, authors should provide estimates of the frequency of the main severe adverse events and reasons for treatment discontinuation separately for each intervention group. If participants may experience an adverse event more than once, the data presented should refer to numbers of affected participants; numbers of adverse events may also be of interest. Authors should provide operational definitions for their measures of the severity of adverse events (174).

Many reports of RCTs provide inadequate information on adverse events. In 192 reports of drug trials, only 39% had adequate reporting of clinical adverse events and 29% had adequate reporting of laboratory-determined toxicity (174). Furthermore, in one volume of a prominent general medical journal in 1998, 58% (30 of 52) of reports (mostly RCTs) did not provide any details on harmful consequences of the interventions (Hasford J. Personal communication).

### Discussion

*Item 20.* Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes.

### Explanation

It has been argued that the discussion sections of scientific reports are filled with rhetoric supporting the authors' findings (175) and provide little measured argument of the pros and cons of the study and its results. Some journals have attempted to remedy this problem by encouraging more structure to authors' discussion of their results (176, 177). For example, *Annals of Internal Medicine* (176) recommends that authors structure the discussion section by presenting 1) a brief synopsis of

the key findings; 2) consideration of possible mechanisms and explanations; 3) comparison with relevant findings from other published studies (whenever possible including a systematic review combining the results of the current study with the results of all previous relevant studies); 4) limitations of the present study (and methods used to minimize and compensate for those limitations); and 5) a brief section that summarizes the clinical and research implications of the work, as appropriate. We recommend that authors follow these sensible suggestions, perhaps also using suitable subheadings in the discussion section.

Although discussion of limitations is frequently omitted from reports of original clinical research (178), identification and discussion of the weaknesses of a study have particular importance. For example, a surgical group recently reported that laparoscopic cholecystectomy, a technically difficult procedure, had significantly lower rates of complications (primary outcome) than the more traditional open cholecystectomy for management of acute cholecystitis (179). However, the authors failed to discuss the potential bias of their results: namely, that the study investigators themselves had completed all the laparoscopic cholecystectomies, whereas 80% of the open cholecystectomies had been completed by trainees. The positive results observed for laparoscopic cholecystectomy may have been merely a function of surgical experience, thus biasing the results. Evaluation of the results in light of this methodologic weakness would have been helpful to readers.

Authors should also discuss any imprecision\* of the results, perhaps when discussing study weaknesses. Imprecision may arise in connection with several aspects of a study, including measurement of a primary outcome (see item 6) or diagnosis (see item 3a). Perhaps the scale used was validated on an adult population but used in a pediatric one, or the assessor was not trained in how to administer the instrument. Issues such as these can lead to imprecise results and should be discussed by the authors.

The difference between statistical significance and clinical importance should always be borne in mind. Authors should particularly avoid the common error of interpreting a nonsignificant result as indicating equivalence of interventions. The confidence interval (item 17) provides valuable insight into whether the trial result is

compatible with a clinically important effect, regardless of the *P* value (94).

Authors should exercise special care when evaluating the results of trials with multiple comparisons\*. Such multiplicity arises from several interventions, outcome measures, time points, subgroup analyses, and other factors. In such circumstances, some statistically significant findings are likely to result from chance alone.

*Item 21.* Generalizability (external validity) of the trial findings.

#### *Example*

Despite the size and duration of this trial, the populations of patients with OA [osteoarthritis] and RA [rheumatoid arthritis] are much larger and therapy continues for substantially longer than 6 months. Moreover, many patients with OA and RA have comorbid illnesses (e.g., active GI [gastrointestinal] disease) that would have excluded them from the current study. Consequently, the results of this study do not address the occurrence of rare adverse events, nor can they be extrapolated to all patients seen in general clinical practice (180).

#### *Explanation*

External validity, also called generalizability or applicability, is the extent to which the results of a study can be generalized to other circumstances (181). Internal validity is a prerequisite for external validity: the results of a flawed trial are invalid and the question of its external validity becomes irrelevant. There is no external validity per se; the term is meaningful only with regard to clearly specified conditions that were not directly examined in the trial. Can results be generalized to an individual patient or groups that differ from those enrolled in the trial with regard to age, sex, severity of disease, and comorbid conditions? Are the results applicable to other drugs within a class of similar drugs, to a different dosage, timing, and route of administration, and to different concomitant therapies? Can the same results be expected at the primary, secondary, and tertiary levels of care? What about the effect on related outcomes that were not assessed in the trial, and the importance of length of follow-up and duration of treatment?

External validity is a matter of judgment and depends on the characteristics of the participants included

in the trial, the trial setting, the treatment regimens tested, and the outcomes assessed (182). It is therefore crucial that adequate information be provided about eligibility criteria and the setting and location (item 3), the interventions and how they were administered (item 4), the definition of outcomes (item 6), and the period of recruitment and follow-up (item 14). The proportion of control group participants in whom the outcome develops (control group risk) is also important.

Several considerations are important when results of a trial are applied to an individual patient (183–185). Although some variation in treatment response between an individual patient and the patients in a trial or systematic review is to be expected, the differences tend to be quantitative rather than qualitative. Although there are important exceptions (185), therapies found to be beneficial in a narrow range of patients generally have broader application in actual practice. Measures that incorporate baseline risk and therapeutic effects, such as the number needed to treat to obtain one additional favorable outcome and the number needed to treat to produce one adverse effect, are helpful in assessing the benefit-to-risk ratio in an individual patient or group with characteristics that differ from the typical trial participant (185–187). Finally, after deriving patient-centered estimates for the potential benefit and harm from an intervention, the clinician must integrate them with the patient's values and preferences for therapy. Similar considerations apply when assessing the generalizability of results to different settings and interventions.

*Item 22.* General interpretation of the results in the context of current evidence.

#### Example

Studies published before 1990 suggested that prophylactic immunotherapy also reduced nosocomial infections in very-low-birth-weight infants. However, these studies enrolled small numbers of patients; employed varied designs, preparations, and doses; and included diverse study populations. In this large multicenter, randomized controlled trial, the repeated prophylactic administration of intravenous immune globulin failed to reduce the incidence of nosocomial infections significantly in premature infants weighing 501 to 1500 g at birth (188).

#### Explanation

The result of an RCT is important regardless of which treatment appears better, magnitude of effect, or precision. Readers will want to know how the present trial's results relate to those of other published RCTs. Ideally, this can be achieved by including a formal systematic review (meta-analysis) in the results or discussion section of the report (82, 189, 190). Such synthesis is relevant only when previous trial results already exist (for example, in the Cochrane Controlled Trials Register [191]) and may often be impractical.

Incorporating a systematic review into the discussion section of a trial report lets the reader interpret the results of the trial as it relates to the totality of evidence. Such information may help readers assess whether the results of the RCT are similar to those of other trials in the same topic area. It may also provide valuable information about the degree of similarity of participants across studies. Recent evidence suggests that reports of RCTs have not adequately dealt with this point (192). Bayesian methods can be used to statistically combine the trial data with previous evidence (193).

We recommend that at a minimum, authors should discuss the results of their trial in the context of existing evidence. This discussion should be as systematic as possible and not limited to studies that support the results of the current trial (194). Ideally, we recommend a systematic review and indication of the potential limitation of the discussion if this cannot be completed.

#### COMMENTS

Assessment of health care interventions can be misleading unless investigators ensure unbiased comparisons. Random allocation to study groups remains the only method that eliminates selection and confounding biases. Indeed, methodologic investigators often (195–197) but not always (198, 199) detect consistent differences when they compare nonrandomized and randomized studies.

Bias jeopardizes even RCTs, however, if investigators carry out such trials improperly (200). Recent results provide empirical evidence that some RCTs have biased results. Four separate studies have found that trials that used inadequate or unclear allocation concealment, compared with those that used adequate concealment, yielded 30% to 40% larger estimates of effect on

average (2, 4, 5, 201). The poorly executed trials tended to exaggerate treatment effects and to have important biases.

Only high-quality research, in which proper attention has been given to design, will consistently eliminate bias. The design and implementation of an RCT require methodologic as well as clinical expertise; meticulous effort (22, 106); and a high index of suspicion for unanticipated difficulties, potentially unnoticed problems, and methodologic deficiencies. Reports of RCTs should be written with similarly close attention to minimizing bias. Readers should not have to speculate; the methods used should be transparent, so that readers can readily differentiate trials with unbiased results from those with questionable results. Sound science encompasses adequate reporting, and the conduct of ethical trials rests on the footing of sound science (202).

We wrote this explanatory article to assist authors in using CONSORT and to explain in general the importance of adequately reporting trials. The CONSORT statement can help researchers designing trials in future and can guide peer reviewers and editors in their evaluation of manuscripts. Because CONSORT is an evolving document, it requires a dynamic process of continual assessment, refinement, and, if necessary, change. Thus, the principles presented in this article and the CONSORT checklist (56–58) are open to change as new evidence and critical comments accumulate.

The first version of the CONSORT statement, despite its limitations, appears to have led to some improvement the quality of reporting of RCTs in the journals that have adopted it (54, 56–58). Other groups are using the CONSORT template to improve the reporting of other research designs, such as diagnostic tests (Lijmer J. Personal communication), meta-analysis of RCTs (203), and meta-analyses of observational studies (204). We hope that this collaborative spirit will continue.

The CONSORT Web site (<http://www.consort-statement.org>) has been established to provide educational material and a repository database of materials relevant to the reporting of RCTs. The site will include many examples from real trials, including all of the examples included in this article. We will continue to add good and weaker examples of reporting to the database, and we invite readers to submit further suggestions to the CONSORT coordinator (Leah Lepage; e-mail,

llepage@uottawa.ca). We will endeavor to make the examples easy to access and disseminate to improve the training of clinical trialists now and in the future.

The CONSORT statement will need periodic re-evaluation until we have direct evidence of the importance of each item on the checklist and flow diagram. The CONSORT group will continue to survey the literature to find relevant articles that address issues to enhance the quality of reporting of RCTs, and we invite authors of any such articles to notify the CONSORT coordinator about them. All of this information will be made accessible through the CONSORT Web site, which will be updated regularly.

The efforts of the CONSORT group have been noticed. Many journals, including *The Lancet*, *British Medical Journal*, *Journal of the American Medical Association*, and *Annals of Internal Medicine*, and a growing number of biomedical editorial groups, including the International Committee of Medical Journals Editors (Vancouver Group) and the Council of Science Editors, have given their official support to CONSORT. We invite other journals concerned about the quality of reporting of clinical trials to adopt the CONSORT statement and contact us through our Web site to let us know of their support. The ultimate benefactors of these collective efforts should be people who, for whatever reason, require intervention from the health care community.

## GLOSSARY

*Adjusted analysis:* Usually refers to attempts to control (adjust) for baseline imbalances between groups in important patient characteristics. Sometimes used to refer to adjustments of *P* value to take account of multiple testing. See *Multiple comparisons*.

*Adverse event:* An unwanted effect detected in participants in a trial. The term is used regardless of whether the effect can be attributed to the intervention under evaluation. See also *Side effect*.

*Allocation concealment:* A technique used to prevent selection bias by concealing the allocation sequence from those assigning participants to intervention groups, until the moment of assignment. Allocation concealment prevents researchers from (unconsciously or otherwise) influencing which participants are assigned to a given intervention group.

*Allocation ratio:* The ratio of intended numbers of participants in each of the comparison groups. For two-group trials, the allocation ratio is usually 1:1, but unequal allocation (such as 1:2) is sometimes used.

*Allocation sequence:* A list of interventions, randomly or-

dered, used to assign sequentially enrolled participants to intervention groups. Also termed “assignment schedule,” “randomization schedule,” or “randomization list.”

*Ascertainment bias:* Systematic distortion of the results of a randomized trial that occurs when the person assessing outcome, whether an investigator or the participant, knows the group assignment.

*Assignment:* See *Random assignment*.

*Baseline characteristics:* Demographic, clinical, and other data collected for each participant at the beginning of the trial, before the intervention is administered. See also *Prognostic variable*.

*Bias:* Systematic distortion of the estimated intervention effect away from the “truth,” caused by inadequacies in the design, conduct, or analysis of a trial.

*Blinding (masking):* The practice of keeping the trial participants, care providers, data collectors, and sometimes those analyzing data unaware of which intervention is being administered to which participant. Blinding is intended to prevent bias on the part of study personnel. The most common application is *double-blinding*, in which participants, caregivers, and outcome assessors are blinded to intervention assignment. The term *masking* may be used instead of blinding.

*Block randomization:* See *Permuted block design*.

*Blocking:* See *Permuted block design*.

*Comparison groups:* The groups being compared in the randomized trial. Also referred to as “study groups”; “treatment groups”; “arms” of a trial; or by individual terms, such as “treatment group” and “control group.”

*Concealment:* See *Allocation concealment*.

*Confidence interval:* A measure of the precision of an estimated value. The interval represents the range of values, consistent with the data, that is believed to encompass the “true” value with high probability (usually 95%). The confidence interval is expressed in the same units as the estimate. Wider intervals indicate lower precision; narrow intervals indicate greater precision.

*Confounding:* A situation in which the estimated intervention effect is biased because of some difference between the comparison groups apart from the planned interventions, such as baseline characteristics, prognostic factors, or concomitant interventions. For a factor to be a confounder, it must differ between the comparison groups and predict the outcome of interest. See also *Adjusted analysis*.

*Deterministic method of allocation:* A method of allocating participants to interventions that uses a predetermined rule without a random element (for example, alternate assignment or based on day of week, hospital number, or date of birth). Because group assignments can be predicted in advance of assignment in deterministic methods, participant allocation may be

manipulated, causing selection bias. See also *Selection bias*; *Allocation concealment*.

*Effect size:* See *Treatment effect*.

*Eligibility criteria:* The clinical and demographic characteristics that define which persons are eligible to be enrolled in a trial.

*End point:* See *Outcome measure*.

*Enrollment:* The act of admitting a participant into a trial. Participants should be enrolled only after study personnel have confirmed that all the eligibility criteria have been met. Formal enrollment must occur before random assignment is performed.

*External validity:* The extent to which the results of a trial provide a correct basis for generalizations to other circumstances. Also called *generalizability* or *applicability*.

*Follow-up:* A process of periodic contact with participants enrolled in the randomized trial for the purpose of administering the assigned interventions, modifying the course of interventions, observing the effects of the interventions, or collecting data. See also *Loss to follow-up*.

*Generation of allocation sequence:* The procedure used to create the (random) sequence for making intervention assignments, such as a table of random numbers or a computerized random-number generator. Such options as simple randomization, blocked randomization, and stratified randomization are part of the generation of the allocation sequence.

*Hypothesis:* In a trial, a statement relating to the possible different effect of the interventions on an outcome. The null hypothesis of no such effect is amenable to explicit statistical evaluation by a hypothesis test, which generates a *P* value.

*Imprecision:* A quantification of the uncertainty in an estimate such as an effect size, usually expressed as the 95% confidence interval around the estimate. Also refers more generally to other sources of uncertainty, such as measurement error.

*Intention-to-treat analysis:* A strategy for analyzing data in which all participants are included in the group to which they were assigned, regardless of whether they completed the intervention given to the group. Intention-to-treat analysis prevents bias caused by loss of participants, which may disrupt the baseline equivalence established by random assignment and may reflect nonadherence to the protocol.

*Interaction:* A situation in which the effect of one explanatory variable on the outcome is affected by the value of a second explanatory variable. In a trial, a test of interaction examines whether the treatment effect varies across subgroups of participants. See also *Subgroup analysis*.

*Interim analysis:* Analysis comparing intervention groups at any time before formal completion of the trial, usually before recruitment is complete. Often used with stopping rules so that a trial can be stopped if participants are being put at risk unne-

sarily. The timing and frequency of interim analyses should be specified in the original trial protocol.

*Internal validity:* The extent to which the design and conduct of the trial eliminate the possibility of bias.

*Intervention:* The treatment or other health care course of action under investigation. The effects of an intervention are quantified by the outcome measures.

*Loss to follow-up:* Loss of contact with some participants, so that researchers cannot complete data collection as planned. Loss to follow-up is a common cause of missing data, especially in long-term studies. See also *Follow-up*.

*Minimization:* An assignment strategy, similar in intention to stratification, that ensures excellent balance between intervention groups for specified prognostic factors. The next participant is assigned to whichever group would minimize the imbalance between groups on specified prognostic factors. Minimization is an acceptable alternative to random assignment (Table 3).

*Multiple comparisons:* Performance of multiple analyses on the same data. Multiple statistical comparisons increase the probability of a type I error: that is, attributing a difference to an intervention when chance is the more likely explanation.

*Multiplicity:* The proliferation of possible comparisons in a trial. Common sources of multiplicity are multiple outcome measures, outcomes assessed at several time points after the intervention, subgroup analyses, or multiple intervention groups.

*Objectives:* The general questions that the trial was designed to answer. The objective may be associated with one or more hypotheses that, when tested, will help answer the question. See also *Hypothesis*.

*Open trial:* A randomized trial in which no one is blinded to group assignment.

*Outcome measure:* An outcome variable of interest in the trial (also called an *end point*). Differences between groups in outcome variables are believed to be the result of the differing interventions. The primary outcome is the outcome of greatest importance. Data on secondary outcomes are used to evaluate additional effects of the intervention.

*Participant:* A person who takes part in a trial. Participants usually must meet certain eligibility criteria. See also *Recruitment, Enrollment*.

*Performance bias:* Systematic differences in the care provided to the participants in the comparison groups other than the intervention under investigation.

*Permuted block design:* An approach to generating an allocation sequence in which the number of assignments to intervention groups satisfies a specified allocation ratio (such as 1:1 or 2:1) after every “block” of specified size. For example, a block size of 12 may contain 6 of A and 6 of B (ratio of 1:1) or 8 of A and 4 of B (ratio of 2:1). Generating the allocation sequence involves

random selection from all of the permutations of assignments that meet the specified ratio.

*Planned analyses:* The statistical analyses specified in the trial protocol (that is, planned in advance of data collection). Also called *a priori analyses*. In contrast to unplanned analyses (also called *exploratory, data-derived, or post hoc analyses*), which are analyses suggested by the data. See also *Subgroup analyses*.

*Power:* The probability (generally calculated before the start of the trial) that a trial will detect as statistically significant an intervention effect of a specified size. The prespecified trial size is often chosen to give the trial the desired power. See *Sample size*.

*Precision:* See *Imprecision*.

*Prognostic variable:* A baseline variable that is prognostic in the absence of intervention. Unrestricted, simple randomization can lead to chance baseline imbalance in prognostic variables, which can affect the results and weaken the trial’s credibility. Stratification and minimization protect against such imbalances. See also *Adjusted analysis, Restricted randomization*.

*Protocol deviation:* A failure to adhere to the prespecified trial protocol, or a participant for whom this occurred. Examples are ineligible participants who were included in the trial by mistake and those for whom the intervention or other procedure differed from that outlined in the protocol.

*Random allocation; random assignment; randomization:* In a randomized trial, the process of assigning participants to groups such that each participant has a known and usually an equal chance of being assigned to a given group. It is intended to ensure that the group assignment cannot be predicted.

*Recruitment:* The process of getting participants into a randomized trial. See also *Enrollment*.

*Restricted randomization:* Any procedure used with random assignment to achieve balance between study groups in size or baseline characteristics. *Blocking* is used to ensure that comparison groups will be of approximately the same size. With *stratification*, randomization with restriction is carried out separately within each of two or more subsets of participants (for example, defining disease severity or study centers) to ensure that the patient characteristics are closely balanced within each intervention group (Table 3).

*Sample size:* The number of participants in the trial. The intended sample size is the number of participants planned to be included in the trial, usually determined by using a statistical power calculation. The sample size should be adequate to provide a high probability of detecting as significant an effect size of a given magnitude if such an effect actually exists. The achieved sample size is the number of participants enrolled, treated, or analyzed in the study.

*Selection bias:* Systematic error in creating intervention groups, causing them to differ with respect to prognosis. That is, the groups differ in measured or unmeasured baseline character-

istics because of the way in which participants were selected for the study or assigned to their study groups. The term is also used to mean that the participants are not representative of the population of all possible participants. See also *Allocation concealment*, *External validity*.

*Side effect*: An unintended, unexpected, or undesirable result of an intervention. See also *Adverse event*.

*Simple randomization*: Randomization without restriction. In a two-group trial, it is analogous to the toss of a coin. See *Restricted randomization*.

*Stopping rule*: In some trials, a statistical criterion that, when met by the accumulating data, indicates that the trial can or should be stopped early to avoid putting participants at risk unnecessarily or because the intervention effect is so great that further data collection is unnecessary. Usually defined in the trial protocol and implemented during a planned interim analysis. See also *Interim analysis*.

*Stratified randomization*: Random assignment within groups defined by participant characteristics, such as age or disease severity, intended to ensure good balance of these factors across intervention groups. See also *Restricted randomization*.

*Subgroup analysis*: An analysis in which the intervention effect is evaluated in a defined subset of the participants in the trial, or in complementary subsets, such as by sex or in age categories. Sample sizes in subgroup analyses are often small, and subgroup analyses therefore usually lack statistical power. They are also subject to the multiple comparisons problem. See also *Multiple comparisons*.

*Treatment effect*: A measure of the difference in outcome between intervention groups. Commonly expressed as a risk ratio (relative risk), odds ratio, or risk difference for binary outcomes and as difference in means for continuous outcomes. Often referred to as the “effect size.”

From ICRF Medical Statistics Group and Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, MRC Health Services Research Collaboration, University of Bristol, Bristol, and London School of Hygiene and Tropical Medicine, London, United Kingdom; Family Health International and The School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; Thomas C. Chalmers Centre for Systematic Reviews, Ottawa, Ontario, Canada; American College of Physicians–American Society of Internal Medicine, Philadelphia, Pennsylvania; Nordic Cochrane Centre, Copenhagen, Denmark; and Tom Lang Communications, Lakewood, Ohio.

**Acknowledgments**: The authors thank Leah Lepage for coordinating the activities of the CONSORT group and Margaret J. Sampson, Louise Roy, and Kaitryn Campbell for creating a database of references.

**Grant Support**: Financial support to convene meetings of the CONSORT group was provided in part by Abbott Laboratories, American College of Physicians–American Society of Internal Medicine, Glaxo

Wellcome, *The Lancet*, Merck, the Canadian Institutes for Health Research, National Library of Medicine, and TAP Pharmaceuticals.

**Requests for Single Reprints**: Leah Lepage, PhD, Thomas C. Chalmers Centre for Systematic Reviews, Children’s Hospital of Eastern Ontario Research Institute, Room R235, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada; e-mail, llepage@uottawa.ca.

**Current Author Addresses**: Professor Altman: ICRF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Old Road, Headington, Oxford OX3 7LF, United Kingdom.

Dr. Schulz: Quantitative Sciences, Family Health International, PO Box 13950, Research Triangle Park, NC 27709.

Mr. Moher: Thomas C. Chalmers Center for Systematic Reviews, Children’s Hospital of Eastern Ontario Research Institute, Room R2226, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada.

Dr. Egger: MRC Health Services Research Collaboration, University of Bristol, Canynge Hall, Whiteladies Road, Bristol B58 2PR, United Kingdom.

Dr. Davidoff: *Annals of Internal Medicine*, American College of Physicians–American Society of Internal Medicine, 190 N. Independence Mall West, Philadelphia, PA 19106.

Professor Elbourne: Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom.

Dr. Gøtzsche: Nordic Cochrane Centre, Rigshospitalet, Dept 7112, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark.

Mr. Lang: 13849 Edgewater Drive, Lakewood, OH 44107.

## References

1. Cochrane AL. Effectiveness and Efficiency: Random Reflections on Health Services. Abingdon, UK: The Nuffield Provincial Hospitals Trust; 1972:2.
2. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408-12. [PMID: 0007823387]
3. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *JAMA*. 1998;279:1489-91. [PMID: 0009600488]
4. Kjaergård L, Villumsen J, Gluud C. Quality of randomised clinical trials affects estimates of intervention efficacy [Abstract]. In: Abstracts for Workshops and Scientific Sessions, 7th International Cochrane Colloquium, Rome, Italy, 1999.
5. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ*. [In press].
6. Veldhuyzen van Zanten SJ, Cleary C, Talley NJ, Peterson TC, Nyren O, Bradley LA, et al. Drug treatment of functional dyspepsia: a systematic analysis of trial methodology with recommendations for design of future trials. *Am J Gastroenterol*. 1996;91:660-73. [PMID: 0008677926]
7. Talley NJ, Owen BK, Boyce P, Paterson K. Psychological treatments for irritable bowel syndrome: a critique of controlled treatment trials. *Am J Gastroenterol*. 1996;91:277-83. [PMID: 0008607493]
8. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol*. 2000;136:381-5. [PMID: 0010724201]
9. Kjaergård LL, Nikolova D, Gluud C. Randomized clinical trials in HEP-

- ATOLOGY: predictors of quality. *Hepatology*. 1999;30:1134-8. [PMID: 0010534332]
10. Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA*. 1966;195:1123-8. [PMID: 0005952081]
  11. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J*. 1977;1:85-7. [PMID: 0000832023]
  12. Hall JC, Hill D, Watts JM. Misuse of statistical methods in the Australasian surgical literature. *Aust N Z J Surg*. 1982;52:541-3. [PMID: 0006959608]
  13. Altman DG. Statistics in medical journals. *Stat Med*. 1982;1:59-71. [PMID: 0007187083]
  14. 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;317:426-32. [PMID: 0003614286]
  15. Altman DG. The scandal of poor medical research [Editorial]. *BMJ*. 1994;308:283-4. [PMID: 0008124111]
  16. DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med*. 1982;306:1332-7. [PMID: 0007070458]
  17. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272:122-4. [PMID: 0008015121]
  18. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319:670-4. [PMID: 0010480822]
  19. Nicolucci A, Grilli R, Alexanian AA, Apolone G, Torri V, Liberati A. Quality, evolution, and clinical implications of randomized, controlled trials on the treatment of lung cancer. A lost opportunity for meta-analysis. *JAMA*. 1989;262:2101-7. [PMID: 0002677423]
  20. Sonis J, Joines J. The quality of clinical trials published in *The Journal of Family Practice*, 1974-1991. *J Fam Pract*. 1994;39:225-35. [PMID: 0008077901]
  21. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA*. 1994;272:125-8. [PMID: 0008015122]
  22. Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet*. 1996;348:596-8. [PMID: 0008774577]
  23. Ah-See KW, Molony NC. A qualitative assessment of randomized controlled trials in otolaryngology. *J Laryngol Otol*. 1998;112:460-3. [PMID: 0009747475]
  24. Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke*. 1998;29:2203-10. [PMID: 0009756604]
  25. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet*. 1990;335:149-53. [PMID: 0001967441]
  26. Williams DH, Davis CE. Reporting of assignment methods in clinical trials. *Control Clin Trials*. 1994;15:294-8. [PMID: 0007956269]
  27. Mosteller F, Gilbert JP, McPeck B. Reporting standards and research strategies for controlled trials. Agenda for the editor. *Control Clin Trials*. 1980;1:37-58.
  28. Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials*. 1989;10:31-56. [PMID: 0002702836]
  29. Tyson JE, Furzan JA, Reisch JS, Mize SG. An evaluation of the quality of therapeutic studies in perinatal medicine. *J Pediatr*. 1983;102:10-3. [PMID: 0006848706]
  30. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet*. 1996;347:363-6. [PMID: 0008598702]
  31. Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *JAMA*. 1998;280:247-9. [PMID: 0009676671]
  32. Altman DG. Randomisation [Editorial]. *BMJ*. 1991;302:1481-2. [PMID: 0001855013]
  33. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ*. 1948;2:769-82.
  34. Schulz KF. Randomized controlled trials. *Clin Obstet Gynecol*. 1998;41:245-56. [PMID: 0009646957]
  35. Armitage P. The role of randomization in clinical trials. *Stat Med*. 1982;1:345-52. [PMID: 0007187102]
  36. Greenland S. Randomization, statistics, and causal inference. *Epidemiology*. 1990;1:421-9. [PMID: 0002090279]
  37. Kleijnen J, Gøtzsche P, Kunz RA, Oxman AD, Chalmers I. So what's so special about randomisation. In: Maynard A, Chalmers I, eds. *Non-Random Reflections on Health Services Research: On the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ; 1997.
  38. Chalmers I. Assembling comparison groups to assess the effects of health care. *J R Soc Med*. 1997;90:379-86. [PMID: 0009290419]
  39. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ*. 1998;317:1181-4. [PMID: 0009794850]
  40. A proposal for structured reporting of randomized controlled trials. The Standards of Reporting Trials Group. *JAMA*. 1994;272:1926-31. [PMID: 0007990245]
  41. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med*. 1994;121:894-5. [PMID: 0007978706]
  42. Rennie D. Reporting randomized controlled trials. An experiment and a call for responses from readers [Editorial]. *JAMA*. 1995;273:1054-5. [PMID: 0007897791]
  43. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637-9. [PMID: 0008773637]
  44. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996;276:1339-41. [PMID: 0008861994]
  45. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. The *BMJ* Economic Evaluation Working Party. *BMJ*. 1996;313:275-83. [PMID: 0008704542]
  46. Lang TA, Secic M. *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers*. Philadelphia: American Coll of Physicians; 1997.
  47. Staquet M, Berzon R, Osoba D, Machin D. Guidelines for reporting results of quality of life assessments in clinical trials. *Qual Life Res*. 1996;5:496-502. [PMID: 0008973129]
  48. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement [Editorial]. *BMJ*. 1996;313:570-1. [PMID: 0008806240]
  49. [A standard method for reporting randomized medical scientific research; the 'Consolidation of the standards of reporting trials']. *Ned Tijdschr Geneeskd*. 1998;142:1089-91. [PMID: 0009623225]
  50. Huston P, Hoey J. CMAJ endorses the CONSORT statement. CONSolidation of Standards for Reporting Trials. *CMAJ*. 1996;155:1277-82. [PMID: 0008911294]
  51. Ausejo M, Saenz A, Moher D. [CONSORT: an attempt to improve the quality of publication of clinical trials] [Editorial]. *Aten Primaria*. 1998;21:351-2. [PMID: 0009633133]

52. Davidoff F. News from the International Committee of Medical Journal Editors [Editorial]. *Ann Intern Med.* 2000;133:229-31. [PMID: 0010906840]
53. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. The CONSORT Group. *JAMA.* 2001;285:1992-5.
54. Egger M, Jüni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. The CONSORT Group. *JAMA.* 2001;285:1996-9.
55. Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. Consolidated Standards of Reporting Trials. *JAMA.* 1998;279:1487-9. [PMID: 0009600487]
56. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. The CONSORT Group. *Ann Intern Med.* 2001;134:657-62.
57. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. The CONSORT Group. *JAMA.* 2001;285:1987-91.
58. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. The CONSORT Group. *Lancet.* [In Press].
59. Pocock SJ. *Clinical Trials: A Practical Approach.* Chichester, UK: John Wiley; 1983.
60. Meinert CL. *Clinical Trials: Design, Conduct, and Analysis.* New York: Oxford Univ Pr; 1986.
61. Friedman LM, Furburg CD, DeMets DL. *Fundamentals of Clinical Trials.* 3rd ed. New York: Springer; 1998.
62. Bolliger CT, Zellweger JP, Danielsson T, van Biljon X, Robidou A, Westin A, et al. Smoking reduction with oral nicotine inhalers: double blind, randomised clinical trial of efficacy and safety. *BMJ.* 2000;321:329-33. [PMID: 0010926587]
63. Prasad AS, Fitzgerald JT, Bao B, Beck FW, Chandrasekar PH. Duration of symptoms and plasma cytokine levels in patients with the common cold treated with zinc acetate. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med.* 2000;133:245-52. [PMID: 0010929163]
64. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ.* 1994;309:1286-91. [PMID: 0007718048]
65. Lefebvre C, Clarke M. Identifying randomised trials. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context.* London: BMJ Books; 2001:69-86.
66. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med.* 1990;113:69-76. [PMID: 0002190518]
67. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association.* *CMAJ.* 1994;150:1611-5. [PMID: 0008174031]
68. Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *Journal of Information Science.* 1996;22:349-56.
69. Dammers JW, Veering MM, Vermeulen M. Injection with methylprednisolone proximal to the carpal tunnel: randomised double blind trial. *BMJ.* 1999;319:884-6. [PMID: 0010506042]
70. Sandler AD, Sutton KA, DeWeese J, Girardi MA, Sheppard V, Bodfish JW. Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive developmental disorder. *N Engl J Med.* 1999;341:1801-6. [PMID: 0010588965]
71. World Medical Association declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *JAMA.* 1997;277:925-6. [PMID: 0009062334]
72. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med.* 1992;327:248-54. [PMID: 0001614465]
73. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ.* 1996;313:1390-3. [PMID: 0008956711]
74. Sinei SK, Schulz KF, Lamptey PR, Grimes DA, Mati JK, Rosenthal SM, et al. Preventing IUCD-related pelvic infection: the efficacy of prophylactic doxycycline at insertion. *Br J Obstet Gynaecol.* 1990;97:412-9. [PMID: 0002196934]
75. Rodgers A, MacMahon S. Systematic underestimation of treatment effects as a result of diagnostic test inaccuracy: implications for the interpretation and design of thromboprophylaxis trials. *Thromb Haemost.* 1995;73:167-71. [PMID: 0007792725]
76. Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. National Surgical Adjuvant Breast and Bowel Program. Pediatric Oncology Group. *J Clin Epidemiol.* 1998;51:69-79. [PMID: 0009474067]
77. Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. *Surgery.* 1996;119:466-72. [PMID: 0008644014]
78. Shapiro SH, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epidemiol.* 2000;53:973-9. [PMID: 0011027928]
79. Taylor MA, Reilly D, Llewellyn-Jones RH, McSharry C, Aitchison TC. Randomised controlled trial of homeopathy versus placebo in perennial allergic rhinitis with overview of four trial series. *BMJ.* 2000;321:471-6. [PMID: 0010948025]
80. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet.* 2000;356:385-90. [PMID: 0010972371]
81. Roberts C. The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Stat Med.* 1999;18:2605-15. [PMID: 0010495459]
82. Sadler LC, Davison T, McCowan LM. A randomised controlled trial and meta-analysis of active management of labour. *BJOG.* 2000;107:909-15. [PMID: 0010901564]
83. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires.* 2nd ed. New York: Oxford Univ Pr; 1996.
84. Streiner D, Normand C. *Health Measurement Scales: A Practical Guide to Their Development and Use.* 2nd ed. Oxford: Oxford Univ Pr; 1995.
85. Sanders C, Egger M, Donovan J, Tallon D, Frankel S. Reporting on quality of life in randomised controlled trials: bibliographic study. *BMJ.* 1998;317:1191-4. [PMID: 0009794853]
86. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry.* 2000;176:249-52. [PMID: 0010755072]
87. Jadad AR, Boyle M, Cunningham C, Kim M, Schachar R. Treatment of Attention-Deficit/Hyperactivity Disorder. Evidence Report/Technology Assessment No. 11. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Healthcare Research and Quality; 1999. AHRQ publication no. 00-E005.
88. Schachter HM, Pham B, King J, Langford S, Moher D. The Efficacy and Safety of Methylphenidate in Attention Deficit Disorder: A Systematic Review and Meta-Analysis. Prepared for the Therapeutics Initiative, Vancouver, B.C., and the British Columbia Ministry for Children and Families. 2000.
89. Kahn JO, Cherng DW, Mayer K, Murray H, Lagakos S. Evaluation of HIV-1 immunogen, an immunologic modifier, administered to patients infected with HIV having 300 to 549 × 10<sup>6</sup>/L CD4 cell counts: a randomized controlled

- trial. *JAMA*. 2000;284:2193-202. [PMID: 0011056590]
90. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. *BMJ*. 1998;317:703-13. [PMID: 0009732337]
91. Heit JA, Elliott CG, Trowbridge AA, Morrey BF, Gent M, Hirsh J. Ardeparin sodium for extended out-of-hospital prophylaxis against venous thromboembolism after total hip or knee replacement. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 2000;132:853-61. [PMID: 0010836911]
92. Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A. Costs and effectiveness of community postnatal support workers: randomised controlled trial. *BMJ*. 2000;321:593-8. [PMID: 0010977833]
93. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ*. 1995;311:1145-8. [PMID: 0007580713]
94. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311:485. [PMID: 0007647644]
95. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med*. 1978;299:690-4. [PMID: 0000355881]
96. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med*. 1984;3:409-22. [PMID: 0006528136]
97. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*. 1994;121:200-6. [PMID: 0008017747]
98. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet*. 1991;338:131-7. [PMID: 0001677062]
99. Galgiani JN, Catanzaro A, Cloud GA, Johnson RH, Williams PL, Mirels LF, et al. Comparison of oral fluconazole and itraconazole for progressive, non-meningeal coccidioidomycosis. a randomized, double-blind trial. *Ann Intern Med*. 2000;133:676-86. [PMID: 0011074900]
100. Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics*. 1987;43:213-23. [PMID: 0003567306]
101. Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med*. 1985;4:521-6. [PMID: 0004089353]
102. Pocock SJ. When to stop a clinical trial. *BMJ*. 1992;305:235-40. [PMID: 0001392832]
103. DeMets DL, Pocock SJ, Julian DG. The agonising negative trend in monitoring of clinical trials. *Lancet*. 1999;354:1983-8. [PMID: 0010622312]
104. Buyse M. Interim analyses, stopping rules and data monitoring in clinical trials in Europe. *Stat Med*. 1993;12:509-20. [PMID: 0008493429]
105. Altman DG, Bland JM. How to randomise. *BMJ*. 1999;319:703-4. [PMID: 0010480833]
106. Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995;274:1456-8. [PMID: 0007474192]
107. Enas GG, Enas NH, Spradlin CT, Wilson MG, Wiltse CG. Baseline comparability in clinical trials. *Drug Information Journal*. 1990;24:541-8.
108. Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does [Editorial]. *BMJ*. 1998;317:362-3. [PMID: 0009694748]
109. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol*. 1999;52:19-26. [PMID: 0009973070]
110. Tu D, Shalay K, Pater J. Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. *Drug Information Journal*. 2000;34:511-23.
111. Chappell LC, Seed PT, Briley AL, Kelly FJ, Lee R, Hunt BJ, et al. Effect of antioxidants on the occurrence of pre-eclampsia in women at increased risk: a randomised trial. *Lancet*. 1999;354:810-6. [PMID: 0010485722]
112. Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med*. 1987;6:315-28. [PMID: 0002887023]
113. Pocock SJ. Statistical aspects of clinical trial design. *Statistician*. 1982;31:1-18.
114. Haag U. Technologies for automating randomized treatment assignment in clinical trials. *Drug Information Journal*. 1998;32:11
115. LaCroix AZ, Ott SM, Ichikawa L, Scholes D, Barlow WE. Low-dose hydrochlorothiazide and preservation of bone mineral density in older adults. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 2000;133:516-26. [PMID: 0011015164]
116. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology*. 1994;44:16-20. [PMID: 0008290055]
117. Gøtzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials*. 1996;17:285-90; discussion 290-3. [PMID: 0008889343]
118. Carley SD, Libetta C, Flavin B, Butler J, Tong N, Sammy I. An open prospective randomised trial to reduce the pain of blood glucose testing: ear versus thumb. *BMJ*. 2000;321:20. [PMID: 0010875827]
119. Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ*. 2000;321:504. [PMID: 0010948038]
120. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001;285:2000-3.
121. Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ*. 1996;312:742-4. [PMID: 0008605459]
122. Cheng K, Smyth RL, Motley J, O'Hea U, Ashby D. Randomized controlled trials in cystic fibrosis (1966-1997) categorized by time, design, and intervention. *Pediatr Pulmonol*. 2000;29:1-7. [PMID: 0010613779]
123. Lang T. Masking or blinding? An unscientific survey of mostly medical journal editors on the great debate. *MedGenMed*. 2000;E25. [PMID: 0011104471]
124. Lao L, Bergman S, Hamilton GR, Langenberg P, Berman B. Evaluation of acupuncture for pain control after oral surgery: a placebo-controlled trial. *Arch Otolaryngol Head Neck Surg*. 1999;125:567-72. [PMID: 0010326816]
125. Quitkin FM, Rabkin JG, Gerald J, Davis JM, Klein DF. Validity of clinical trials of antidepressants. *Am J Psychiatry*. 2000;157:327-37. [PMID: 0010698806]
126. Nacul LC, Kirkwood BR, Arthur P, Morris SS, Magalhaes M, Fink MC. Randomised, double blind, placebo controlled clinical trial of efficacy of vitamin A treatment in non-measles childhood pneumonia. *BMJ*. 1997;315:505-10. [PMID: 0009329303]
127. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. 2nd ed. London: BMJ Books; 2000:171-90.
128. Altman DG, Bland JM. Statistics notes. Units of analysis. *BMJ*. 1997;314:1874. [PMID: 0009224131]
129. Bolton S. Independence and statistical inference in clinical trial designs: a tutorial review. *J Clin Pharmacol*. 1998;38:408-12. [PMID: 0009602951]
130. Greenland S. Principles of multilevel modelling. *Int J Epidemiol*. 2000;29:158-67. [PMID: 0010750618]

131. Saunders M, Dische S, Barrett A, Harvey A, Gibson D, Parmar M. Continuous hyperfractionated accelerated radiotherapy (CHART) versus conventional radiotherapy in non-small-cell lung cancer: a randomised multicentre trial. CHART Steering Committee. *Lancet*. 1997;350:161-5. [PMID: 0009250182]
132. Matthews JN, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ*. 1996;313:862. [PMID: 0008870577]
133. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355:1064-9. [PMID: 0010744093]
134. Matthews JN, Altman DG. Statistics notes. Interaction 2: Compare effect sizes not P values. *BMJ*. 1996;313:808. [PMID: 0008842080]
135. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116:78-84. [PMID: 0001530753]
136. Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *Am Heart J*. 2000;139:745-51. [PMID: 0010783203]
137. Altman DG. Adjustment for covariate imbalance. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. Chichester, UK: John Wiley; 1998:1000-5.
138. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med*. 1993;118:201-10. [PMID: 0008417638]
139. Bender R, Grouven U. Logistic regression models used in medical research are poorly presented [Letter]. *BMJ*. 1996;313:628. [PMID: 0008806274]
140. Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology literature. *Obstet Gynecol*. 1999;93:1014-20. [PMID: 0010362173]
141. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med*. 1979;301:1410-2.
142. May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation*. 1981;64:669-73. [PMID: 0007023743]
143. Altman DG, Cuzick J, Peto J. More on zidovudine in asymptomatic HIV infection [Letter]. *N Engl J Med*. 1994;330:1758-9. [PMID: 0008190146]
144. Serruys PW, van Hout B, Bonnier H, Legrand V, Garcia E, Macaya C, et al. Randomised comparison of implantation of heparin-coated stents with balloon angioplasty in selected patients with coronary artery disease (Benestent II). *Lancet*. 1998;352:673-81. [PMID: 0009728982]
145. Bove G, Nilsson N. Spinal manipulation in the treatment of episodic tension-type headache: a randomized controlled trial. *JAMA*. 1998;280:1576-9. [PMID: 0009820258]
146. Cunningham D, Pyrhönen S, James RD, Punt CJ, Hickish TF, Heikkilä R, et al. Randomised trial of irinotecan plus supportive care versus supportive care alone after fluorouracil failure for patients with metastatic colorectal cancer. *Lancet*. 1998;352:1413-8. [PMID: 0009807987]
147. van Loon AJ, Mantingh A, Serlier EK, Kroon G, Mooyaart EL, Huisjes HJ. Randomised controlled trial of magnetic-resonance pelvimetry in breech presentation at term. *Lancet*. 1997;350:1799-804. [PMID: 0009428250]
148. Brown MJ, Palmer CR, Castaigne A, de Leeuw PW, Mancía G, Rosenthal T, et al. Morbidity and mortality in patients randomised to double-blind treatment with a long-acting calcium-channel blocker or diuretic in the International Nifedipine GITS study: Intervention as a Goal in Hypertension Treatment (INSIGHT). *Lancet*. 2000;356:366-72. [PMID: 0010972368]
149. Shuster JJ. Median follow-up in clinical trials [Letter]. *J Clin Oncol*. 1991;9:191-2. [PMID: 0001985169]
150. Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72:511-8. [PMID: 0007640241]
151. Senn S. Base logic: tests of baseline balance in randomized clinical trials. *Clinical Research and Regulatory Affairs*. 1995;12:171-82.
152. Haderslev KV, Tjellesen L, Sorensen HA, Staun M. Alendronate increases lumbar spine bone mineral density in patients with Crohn's disease. *Gastroenterology*. 2000;119:639-46. [PMID: 0010982756]
153. Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA*. 1970;211:1993-2003. [PMID: 0005467158]
154. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med*. 1991;10:1595-605. [PMID: 0001947515]
155. Lewis JA, Machin D. Intention to treat—who should use ITT? [Editorial] *Br J Cancer*. 1993;68:647-50. [PMID: 0008398686]
156. Lachin JL. Statistical considerations in the intent-to-treat principle. *Control Clin Trials*. 2000;21:526. [PMID: 0011018568]
157. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther*. 1995;57:6-15. [PMID: 0007828382]
158. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med*. 2000;19:1849-64. [PMID: 0010867675]
159. Ruiz-Canela M, Martínez-González MA, de Irala-Estévez J. Intention to treat analysis is related to methodological quality [Letter]. *BMJ*. 2000;320:1007-8. [PMID: 0010753165]
160. Altman DG. Confidence intervals in practice. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. 2nd ed. London: BMJ Books; 2000:6-14.
161. Altman DG. Clinical trials and meta-analyses. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. 2nd ed. London: BMJ Books; 2000:120-38.
162. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *Ann Intern Med*. 1997;126:36-47. [PMID: 0008992922]
163. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)*. 1986;292:746-50. [PMID: 0003082422]
164. Bailar JC 3rd, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Ann Intern Med*. 1988;108:266-73. [PMID: 0003341656]
165. Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistics*. 2000;49:359-70.
166. Egger M, Dickersin K, Davey Smith G. Problems and limitations in conducting systematic reviews. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books; 2001:43-68.
167. Bland JM. Quoting intermediate analyses can only mislead [Letter]. *BMJ*. 1997;314:1907-8. [PMID: 0009224157]
168. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ*. 1995;310:452-4. [PMID: 0007873954]
169. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ*. 1999;319:1492-5. [PMID: 0010582940]
170. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science*. 1977;198:679-84. [PMID: 0000333584]
171. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266:93-8. [PMID: 0002046134]
172. Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses

- within studies. *Stat Med*. 2000;19:3325-3336. [PMID: 0011122498]
173. Levin M, Quint PA, Goldstein B, Barton P, Bradley JS, Shemie SD, et al. Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal sepsis: a randomised trial. rBPI21 Meningococcal Sepsis Study Group. *Lancet*. 2000;356:961-7. [PMID: 0011041396]
174. Ioannidis JP, Lan J. Completeness of safety reporting in randomized trials. An evaluation of 7 medical areas. *JAMA*. 2001;285:437-43. [PMID: 0011242428]
175. Horton R. The rhetoric of research. *BMJ*. 1995;310:985-7. [PMID: 0007728037]
176. *Annals of Internal Medicine*. Information for authors. Available at [www.annals.org](http://www.annals.org). Accessed 10 January 2001.
177. Docherty M, Smith R. The case for structuring the discussion of scientific papers [Editorial]. *BMJ*. 1999;318:1224-5. [PMID: 0010231230]
178. Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *JAMA*. 1998;280:227-8. [PMID: 0009676663]
179. Kiviluoto T, Sirén J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet*. 1998;351:321-5. [PMID: 0009652612]
180. Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, et al. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: A randomized controlled trial. Celecoxib Long-term Arthritis Safety Study. *JAMA*. 2000;284:1247-55. [PMID: 0010979111]
181. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull*. 1957;54:297-312.
182. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books; 2001.
183. Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA*. 1998;279:545-9. [PMID: 0009480367]
184. Davey Smith G, Egger M. Who benefits from medical interventions? [Editorial] *BMJ*. 1994;308:72-4. [PMID: 0008298415]
185. McAlister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books; 2001.
186. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318:1728-33. [PMID: 0003374545]
187. Altman DG. Confidence intervals for the number needed to treat. *BMJ*. 1998;317:1309-12. [PMID: 0009804726]
188. Fanaroff AA, Korones SB, Wright LL, Wright EC, Poland RL, Bauer CB, et al. A controlled trial of intravenous immune globulin to reduce nosocomial infections in very-low-birth-weight infants. National Institute of Child Health and Human Development Neonatal Research Network. *N Engl J Med*. 1994;330:1107-13. [PMID: 0008133853]
189. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. First International Study of Infarct Survival Collaborative Group. *Lancet*. 1986;2:57-66. [PMID: 0002873379]
190. Gøtzsche PC, Gjørup I, Bonnén H, Brahe NE, Becker U, Burcharth F. Somatostatin v placebo in bleeding oesophageal varices: randomised trial and meta-analysis. *BMJ*. 1995;310:1495-8. [PMID: 0007787594]
191. Cochrane Collaboration. *The Cochrane Library*. Issue 1. Oxford: Update Software; 2000.
192. Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? *JAMA*. 1998;280:280-2. [PMID: 0009676682]
193. Goodman SN. Toward evidence-based medical statistics. 1: The *P* value fallacy. *Ann Intern Med*. 1999;130:995-1004. [PMID: 0010383371]
194. Gøtzsche PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)*. 1987;295:654-6. [PMID: 0003117277]
195. Berlin JA, Begg C, Louis TA. An assessment of publication bias using a sample of published clinical trials. *Journal of the American Statistical Association*. 1989;84:381-92.
196. Guyatt GH, DiCenso A, Farewell V, Willan A, Griffith L. Randomized trials versus observational studies in adolescent pregnancy prevention. *J Clin Epidemiol*. 2000;53:167-74. [PMID: 0010729689]
197. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*. 1998;317:1185-90. [PMID: 0009794851]
198. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342:1878-86. [PMID: 0010861324]
199. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887-92. [PMID: 0010861325]
200. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity. I: clinical trials. *Lancet*. 2001;357:373-80. [PMID: 0011211013]
201. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352:609-13. [PMID: 0009746022]
202. Murray GD. Promoting good research practice. *Stat Methods Med Res*. 2000;9:17-24. [PMID: 0010826155]
203. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet*. 1999;354:1896-900. [PMID: 0010584742]
204. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283:2008-12. [PMID: 0010789670]