A Proposal for Structured Reporting of Randomized Controlled Trials

The Standards of Reporting Trials Group

A RANDOMIZED controlled trial (RCT) is the most reliable method of assessing the efficacy of health care interventions.[1,2] Reports of RCTs should provide readers with adequate information about what went on in the design, execution, analysis, and interpretation of the trial. Such reports will help readers judge the validity of the trial.

There have been several investigations evaluating how RCTs are reported. In an early study, Mahon and Daniel[3] reviewed 203 reports of drug trials published between 1956 and 1960 in the *Canadian Medical Association Journal*. Only 11 reports (5.4%) fulfilled their criteria of a valid report. In a review of 45 trials published during 1985 in three leading general medical journals, Pocock and colleagues[4] reported that a statement about sample size was only mentioned in five (11.1%) of the reports, that only six (13.3%) made use of confidence intervals, and that the statistical analyses tended to exaggerate intervention efficacy. Altman and Doré[5] reviewed 80 reports of trials published in 1987 and 1988 and found that information about the type of randomization was only reported in 32 (40%) of the trials. Schulz and colleagues[6] reviewed 206 reports of trials published during 1990 and 1991 in two British and two US obstetrics and gynecology journals. Only 66 (32%) of the trials reported on how the randomization sequence was generated and only 47 (22.8%) reported on how intervention assignment was concealed until the allocation of therapy. Gøtzsche[7] has suggested the quality of RCT reports in rheumatology may be so weak that it may be impossible to place any confidence in the statistical analyses or conclusions.

A few years ago, to improve the quality of reporting of clinical research, more informative abstracts were developed.[8] Such abstracts provide readers with a series of headings pertaining to the design, conduct, and analysis of a trial and standardized information within each heading. Evidence to date indicates that more informative abstracts have had a positive impact on how the results of abstracts are communicated.[9] More recently, there has been a call to extend more informative abstracts to include structured reporting of the text of each RCT.[10]

A workshop was held in Ottawa, Ontario, on October 7 and 8, 1993, with the aim of developing a new scale to assess the quality of RCT reports. However, during preliminary discussions, participants felt that many of the suggested scale items were irrelevant because they were not regularly reported by authors. There was unanimous agreement that the remainder of the workshop should focus on ways to improve the reporting of RCTs. This report provides a summary of that workshop.

METHODS

Based on prior research,[11] we identified investigators who had published or were developing a scale to assess quality. We also identified researchers who had experience in the design, conduct, and analysis of RCTs, as well as those who had made several contributions to the methodology of RCTs. Our intent was to keep the number of invitees small while comprehensive in scope. Once all potential participants were identified, invitations to participate in the workshop were sent to 32 individuals, of whom 30 agreed to participate.

A 144-item questionnaire was sent to invitees (n=19) primarily responsible for the development of a scale. The questionnaire asked participants to rate which content areas and items they thought important to include when assessing the quality of an RCT. These content areas focused on general characteristics (ie, sample size, informed consent, description of patients, interventions, and outcomes) and internal validity (ie, participant assignment, masking, participant follow-up, and approaches to statistical analysis) of a trial. Based in part on the questionnaire responses, four content areas were identified (participant assignment, masking, participant follow-up, and statistical analysis) for discussion during the workshop.

Once the content areas were decided, all participants were sent a questionnaire asking them to indicate the content area groups in which they would like to participate. Based in part on these responses, participants were divided into one of the groups. We attempted to balance each small group with a scale developer, a clinical trialist, and a content specialist. The workshop comprised small group and plenary sessions. Each small group had an invited chair and a rapporteur, and each plenary session had an invited chair.

The group as a whole was first asked to agree on definitions of "quality" and "structured reporting." Each small group was asked to identify items they thought should be included in a structured report of an RCT. To identify items and decide whether they were considered essential or not, the small groups used a modified Delphi process. It was emphasized that this selection of items should be based on elements of trial design and conduct relevant to internal validity. For each item chosen, the small groups were asked to state whether its importance was based on conviction alone or on empirical research, and failure to clearly report such information was associated with systematically different estimates of intervention effects. For example, Schulz and colleagues[12,13] reviewed 250 reports of RCTs and found the odds ratios in the unclearly concealed trials were on average 30% (95% confidence interval, 21% to 38%) lower than in the adequately concealed trials

A complete list of the Standards of Reporting Trials Group appears at the end of the article.

Reprint requests to the Clinical Epidemiology Unit, Loeb Medical Research Institute, Ottawa Civic Hospital, 1053 Carling Ave, Ottawa, Ontario, Canada K1Y 4E9 (Mr Moher).

(ie, they estimated the intervention to be more effective than it really was). Items not based on empirical evidence were identified to generate a research agenda for the future.

The small groups were also asked to identify which items should be included in a checklist to be used by readers when reviewing an RCT. The chair and rapporteur of each small group made a short presentation of the results of their session at the plenary sessions, and these were followed by discussions with all workshop participants. As this article was going through its revisions, the group reemphasized that this is an evolutionary process. This process will be continued within the Cochrane Collaboration[14] and its Methods Working Groups as well as with others.

RESULTS

We defined structured reporting as "providing sufficiently detailed information about the design, conduct and analysis of the trial for the reader to have confidence that the report is an accurate reflection of what occurred during the various stages of the trial." We defined trial quality as "the confidence that the trial design, conduct, analysis, and presentation has minimized or avoided biases in its intervention comparisons," thus focusing on the internal validity of a trial. Assessing trial quality is based on a desire to estimate the likelihood that the trial results provide a valid (unbiased) estimate of the "truth"—something we cannot observe.

There were 32 items proposed for inclusion in the checklist to be used when preparing a report of an RCT (Table). This article elaborates on 24 items (numbered to correspond to the Table) the group felt to be essential. The results are divided into the four content areas.

Checklist to Be Used by Authors When Preparing or by Readers When Analyzing a Report of a Randomized Controlled Trial

Item	Yes	No	Unable to Determine
1. State the unit of assignment.	☐	☐	☐
2. State the method used to generate the intervention assignment schedule.	☐	☐	☐
3. Describe the method used to conceal the intervention assignment schedule from participants and clinicians until recruitment was complete and irrevocable.	☐	☐	☐
4. Describe the method(s) used to separate the generator and executor of the assignment.	☐	☐	☐
5. Describe an auditable process of executing the assignment method.	☐	☐	☐
6. Identify and compare the distributions of important prognostic characteristics and demographics at baseline.	☐	☐	☐
7. State the method of masking.	☐	☐	☐
8. State how frequently care providers were aware of the intervention allocation, by intervention group.	☐	☐	☐
9. State how frequently participants were aware of the intervention allocation, by intervention group.	☐	☐	☐
10. State whether (and how) outcome assessors were aware of the intervention allocation, by intervention group.	☐	☐	☐
11. State whether the investigator was unaware of trends in the study at the time of participant assignment.	☐	☐	☐
12. State whether masking was successfully achieved for the trial.	☐	☐	☐
13. State whether the data analyst was aware of intervention allocation.*	☐	☐	☐
14. State whether individual participant data were entered into the trial database without awareness of intervention allocation.	☐	☐	☐
15. State whether the data analyst was masked to intervention allocation.	☐	☐	☐
16. Describe fully the numbers and flow of participants, by intervention group, throughout the trial.	☐	☐	☐
17. State clearly the average duration of the trial, by intervention group, and the start and closure dates for the trial.†	☐	☐	☐
18. Report the reason for dropout clearly, by intervention group.	☐	☐	☐
19. Describe the actual timing of measurements, by intervention group.	☐	☐	☐
20. State the predefined primary outcome(s) and analyses clearly.	☐	☐	☐
21. Describe clearly whether the primary analysis has used the intention-to-treat principle.	☐	☐	☐
22. State the intended sample size and its justification.	☐	☐	☐
23. State and explain why the trial is being reported now.	☐	☐	☐
24. Describe and/or compare trial dropouts and completers.	☐	☐	☐
25. State or reference the reliability, validity, and standardization of the primary outcome.‡	☐	☐	☐
26. Define what constituted adverse events and how they were monitored by intervention group.	☐	☐	☐
27. State the appropriate analytical techniques applied to the primary outcome measure(s).	☐	☐	☐
28. Present appropriate measures of variability (eg, confidence intervals for primary outcome measures).	☐	☐	☐
29. Present sufficient simple (unadjusted) summary data on primary outcome measures and important side effects so that the reader can reproduce the results.	☐	☐	☐
30. State the actual probability value and the nature of the significance test.	☐	☐	☐
31. Present appropriate interpretations (eg, NS, no effect; $P<.05$, proof).	☐	☐	☐
32. Present the appropriate emphasis in displaying and interpreting the statistical analysis, in particular controlling for unplanned comparisons.	☐	☐	☐

*If the data analyst is not masked as to the interventions, new treatments may be grossly favored over standard treatments.[6,15]

†This information may sometimes reveal duplicate publication rather than two separate trials by the same author(s).

‡Many trials are longitudinal and require several follow-up assessments. These assessments may be subjective based on the responses of questionnaires or scales. There is wide variation in how scales and questionnaires are constructed,[16,17] which may influence the assessment, reliability, validity, and responsiveness of the treatment outcome of interest. Providing information or references about the development of these outcome measures will enable readers to judge how confident they should be about the results.

CONTENT AREA 1—PARTICIPANT ASSIGNMENT

1. State the Unit of Assignment (eg, Individual, Cluster, or Geographic Region).—If the unit of assignment is not at the individual participant level, but a group, a rationale and description must be provided. Trials that use cluster randomization and are analyzed using standard statistical techniques applied to individual participants can lead, without appropriate adjustment, to invalid results.[18] Appropriate ways of analyzing and reporting group or cluster assignments are provided elsewhere.[19]

2. State the Method Used to Generate the Intervention Assignment Schedule (eg, Random Numbers Table, Computer Random Number Generator, or Random Permuted Blocks Stratified by What Factor).—If random allocation is not used, the alternative method (eg, day of the week) needs to be justified and described. Many trial reports do not provide essential information about assignment.[20] One of the principal advantages of randomization, noted by Hill,[21] is to ensure unpredictability of the intervention group to which the next participant will be assigned. Randomization is also essential to ensure balance for any unknown as well as known confounders, permitting unbiased comparisons between intervention groups.[2,20]

3. Describe the Method Used to Conceal the Intervention Assignment Schedule From Participants and Clinicians Until Recruitment Was Complete and Irrevocable.—To avoid selection bias in participant assignment to intervention, concealment is essential and should be feasible in all trials. Chalmers and colleagues[22] reviewed 145 reports of RCTs in the intervention for acute myocardial infarction to assess whether concealment of participant assignment was associated with estimates of intervention effects. Their results showed that trials that reported concealed assignment yielded smaller differences in case-fatality rates than trials in which assignment was unconcealed (ie, the latter were biased). This empirical work has been supported by Schulz.[12,13]

4. Describe the Method(s) Used to Separate the Generator and Executor of the Assignment.—We have defined

generator as "the individual(s) who, using some bias-free method (eg, central randomization, numbered containers, or sealed opaque envelopes), generated the listing that identified the intervention assignment for every participant," and executor as "the individual(s) who, having determined a participant's eligibility, consults the assignment system for that participant's intervention designation." The person(s) who prepared the randomization scheme ideally should not be involved in determining eligibility, administering intervention, or assessing outcome. That is obviously important because regardless of the methodological quality of the randomization process, such an individual would always have access to the allocation schedule and, thus, the opportunity to introduce bias.

5. Describe an Auditable Process of Executing the Assignment Method.—It should be possible for a review body to reconstruct, from trial records, exactly how each assignment was made. If the process of participant assignment to intervention cannot be reconstructed, the validity of the methodology and, thus, trial results may be in question. Investigators should indicate that adequate trial records have been kept.

6. Identify and Compare the Distributions of Important Prognostic Characteristics and Demographics at Baseline.—We have defined distribution as "the complete summary of the frequencies of the values or categories of a measurement made on a group of persons."[23] When reporting the results of any trial, it is important to present information regarding the comparability of intervention groups. This can best be achieved by presenting data on the distributions of measured baseline variables (other than intervention) thought to affect the outcome. Although *P* values are often used to compare the characteristics of the intervention groups, these statistical comparisons are inappropriate, unless the investigators suspect that the randomization schedule has not been adhered to. A discussion of appropriate ways to present comparability between intervention groups can be found elsewhere.[24]

CONTENT AREA 2—MASKING

7. State the Method of Masking (eg, Physical Characteristics [Whether the Interventions Look and Taste the Same] or Route of Administration).—If masking has not been used, sufficient justification should be provided. Many trial reports do not provide detailed information as to how masking was carried out,[7] whether single, double, or triple, and evidence indicates that masking can affect estimates of intervention effects on subjective outcomes.[12,13,25,26]

8. State How Frequently Care Providers Were Aware of the Intervention Allocation, by Intervention Group.—To reduce bias, it is important that caregivers be unaware of intervention group assignment.[27] For example, contamination bias, in which participants assigned to the control group receive the experimental intervention, and cointervention bias, in which participants assigned to either the control or the experimental group receive additional therapy, can influence intervention results.[28,29] Investigators should also report on how they assessed whether caregivers knew the intervention group to which the participants were being assigned.

9. State How Frequently Participants Were Aware of the Intervention Allocation, by Intervention Group.—To reduce bias, it is important that participants be unaware of their intervention group assignment. Otherwise, participants may become more aware of and may report more symptoms, leading to biased results.[26] Investigators should also report on how they assessed whether participants knew their assignment group.

10. State Whether (and How) Outcome Assessors Were Aware of the Intervention Allocation, by Intervention Group.—We recognize that the outcome assessor as described here may be the caregiver as described in item 7 (indeed, we use the latter only when we cannot assure the former). When study clinicians guess intervention allocations at levels greater than chance, it need not be due to code-breaking, but may merely represent their suspicion about efficacy. For example, before the Canadian aspirin trial,[29] sulfinpyrazone (but not aspirin) was thought to be efficacious for transient ischemic attacks, whereas the trial found that the reverse was true. An end-of-study polling of study clinicians found them to be statistically significantly wrong about intervention allocations; when participants fared well, their clinicians tended to believe that they had been on sulfinpyrazone.

14. State Whether Individual Participant Data Were Entered Into the Trial Database Without Awareness of Intervention Allocation, and 15. Whether Data Analyst Was Masked to Intervention Allocation.—When changes in the measurement scale are made, such as from a continuous to a binomial scale, results can be altered from statistically not significant to significant and vice versa depending on the cutoff points used.[30] Gøtzsche[7] has shown that unmasked data analysis can favor new interventions over standard interventions. To prevent this, cutoff points should be declared a priori, or rationale and cutoff points should be described.

CONTENT AREA 3—PARTICIPANT FOLLOW-UP

16. Describe Fully the Numbers and Flow of Participants, by Intervention Group, Throughout the Trial.—It is important to know how participants were followed up from the time of their randomization until they completed the trial. Unfortunately, it is often difficult to ascertain from reports whether participants were unavailable for follow-up during the course of a trial.[12,13] Without knowledge of this information, readers can make false conclusions about the efficacy of therapy. In a trial that compared medical vs surgical therapy for carotid stenosis,[31] an analysis limited to those participants available for follow-up was statistically in favor of surgical intervention to reduce transient ischemic attacks, stroke, or death. However, when all participants had been accounted for and added to the analysis, the advantage of surgery was no longer statistically significant. Similar, less dramatic examples can be found elsewhere.[32]

Information on the flow of participants should include the number of participants eligible, randomized, treated, completing, and failing to complete the trial, by intervention group. Readers are likely to be in a better position to make decisions if the report documents what happened to all participants from the time they were initially asked to participate to when they were included in the analysis. The flow of participants through a trial can best be represented in a simple flow diagram (Figure).

18. Report the Reason for Dropout Clearly, by Intervention Group.—Participant dropout can be due to a host of factors (such as relocation or adverse reactions), some of which may be related to outcome. Information on dropouts should include reasons for dropout. Providing this information to readers will enable them to judge how confident they should feel about the results.[33] Information about participant dropout can often be best represented in a simple flow diagram (Figure).

19. Describe the Actual Timing of Measurements by Intervention Group.—If intervention groups are followed up with different intensity, this may lead to unmasking, cointervention, and distortion of outcome measurements. This information can best be represented in a simple flow diagram (Figure).

CONTENT AREA 4—APPROACHES TO STATISTICAL ANALYSIS

Methods Section

20. State the Predefined Primary Outcome(s) and Analyses Clearly.—The primary outcome(s), the main event or condition that the trial is designed to evaluate, differs from the secondary out-

come(s), which is considered less important. If the primary outcome is not specified a priori and all outcomes are treated alike, there is an increased risk that multiple analyses will result in false-positive, statistically significant results merely by chance.[34] There is a need for consistency between the predefined primary outcome and the actual primary analysis.

21. Describe Clearly Whether the Primary Analysis Has Used the Intention-to-Treat Principle.—For the majority of trials, the preferred analysis is based on including all participants and their follow-up results in the intervention groups as initially assigned (intention to treat), although there may be important exceptions to this.[27,35] Analysis based on the intervention participants actually received, rather than the intervention to which they were initially assigned (efficacy analysis), may produce invalid results.[36,37] If such efficacy analysis is used, it should be clearly reported and justified.

22. State the Intended Sample Size and Its Justification.—Surveys have shown that most trials do not report their intended sample size.[4,5,7,38,39] Describing and justifying, a priori, the size of a trial gives the reader a clear idea of what potential intervention differences the investigators were interested in detecting. Trials with insufficient sample size may result in potentially useful new therapies being ignored and/or therapies with significant toxicity being accepted. Recommendations on including sample size justification in a structured report have been made elsewhere.[10]

23. State and Explain Why the Trial Is Being Reported Now.—When a trial is being reported, it is important to know whether the trial has gone its intended course or has been terminated early (or late). The reasons for early (or late) termination, such as the discovery of large differences in outcome between intervention groups, unacceptably high adverse events in one group, or slow recruitment, should be specified.[40,41]

Results Section

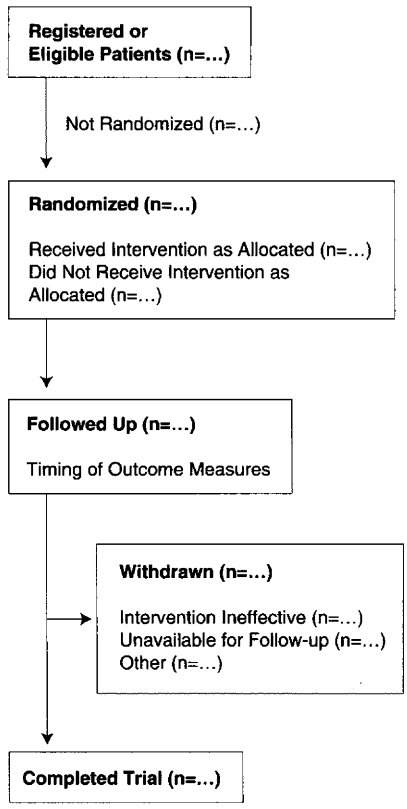
27. State the Appropriate Analytic Techniques Applied to the Primary Outcome Measure(s).—Evans and Pollock[42] reviewed the statistical methods reported in 45 trials in the antibiotic literature. In 31 reports (68.9%), the statistical methods were considered incorrect. In 11 reports (24.4%), the conclusions were not supported by the data. Similarly, in a survey of 20 trials submitted to the *British Medical Journal*, inappropriate analyses were reported in 12 trials (60%).[43] When comparing outcomes between three or more intervention groups, multiple pairwise tests will inflate the overall type I error, increasing the risks of a false-positive result.[34] Several solutions exist, including the following: specifying the primary contrasts beforehand, correcting for multiple comparisons, using another statistical test (such as an analysis of variance), or declaring openly that multiple comparisons have been made. Statistical methods are also likely to be improved by incorporating a statistician in the project,[44] particularly at the beginning.[32]

28. Present Appropriate Measures of Variability (eg, Confidence Intervals) for Primary Outcomes.—Appropriate measures of statistical uncertainty should be reported along with the measure of central tendency. The standard error and confidence intervals are two closely related ways to describe such uncertainty. Presenting confidence intervals for any key estimate enables the reader to see the range within which the true effect or association may plausibly lie, rather than simply assessing whether it is statistically significant.[45]

29. Present Sufficient Simple (Unadjusted) Summary Data on Primary Outcome Measures and Important Side Effects so That the Reader Can Reproduce the Results (eg, Both Numbers of Participants and Numbers of Events, Such as p_1, p_2, n_1, and n_2).—Such presentations allow the reader to make a basic assessment of the intervention(s) and its potential risks.

30. State the Actual Probability Value and the Nature of the Significance Test.—Reporting the actual probability value provides the reader with a precise statement as to the significance of the trial result (eg, P=.02 rather than P<.05; P=.20 rather than "not significant" or NS). Although the probability value P=.06 is not statistically significant at the 5% level and P=.04 is, readers should not interpret these results as substantially different. A statement of the obtained values from the primary statistical test(s) should also be considered when reporting actual probability values (eg, P=.02, z=2.33). It is also important to state (preferably a priori) whether significance testing is one-sided or two-sided. The one-sided approach only assesses whether the experimental therapy is better than the standard under consideration, but because one should always also consider that the experimental intervention may instead be inferior,[46] we recommend use of a two-sided P value. If on rare occasions the one-sided approach is used, it should be described and justified clearly.

There is often particular excitement when a P value falls into the <.05 category, but such an obsession with an arbitrary cutoff is entirely inappropriate. Unfortunately, P values do not express many important attributes of a trial's result: the overall magnitude and consistency of direction in the intervention difference and its range of uncertainty. The confidence interval addresses many of these issues, incorporating statistical as well as clinical significance.[45,47]



Flow diagram of how participants can be represented passing through the various stages of a trial, including withdrawals and timing of outcome measurements.

31. Present Appropriate Interpretations (eg, NS, No Effect; P<.05, Proof).—Trials that report no statistically significant differences between intervention groups may conceal clinically important differences that could not be detected because of small sample size.[39,48] That is, small trials with apparently negative results may be erroneously reported as proof of no intervention benefit. Less commonly, a large trial that reports statistically significant results could have clinical differences that are too small to be of practical importance.[49]

32. Present the Appropriate Emphasis in Displaying and Interpreting the Statistical Analysis, in Particular Controlling for Unplanned Comparisons (eg, Subgroups, Multiple Outcome Measures, or Multiple Analyses).—Many trials report several outcome measures, repeated measures over time or subgroups, without appropriate adjustment for the consequent inflation of the type I error (risk of rejecting the null hypothesis of no

intervention effect when in fact it is true). It is important to report all the challenges to the data, how many times they have been evaluated for statistical significance. If there is no statistical adjustment, this needs to be stated and justified. It is also important to report what methods (if any) have been used to control type I errors.[50]

Subgroup analysis is commonly conducted and inappropriately interpreted in many RCTs. Such analyses are more credible[51] when they have a rationale, are stated a priori, have sufficient statistical power, and are appropriately carried out and interpreted cautiously, with appropriate use of statistical tests of interaction. There is evidence that subgroup analyses can inappropriately affect intervention recommendations.[52] Oxman and Guyatt[52] and Yusuf and colleagues[53] provide valuable guidelines to the analysis and interpretation of subgroup analyses. Statistical presentation is not only providing factual information, but also relying on a judicious selection of appropriate presentation of data.

COMMENT

Clinical trialists have had at least four decades in which to improve the reporting of RCTs. Several guidelines have been published[54-57] to facilitate this process, describing what needs to be included when reporting a trial. In addition, some journals[58] have published checklists of items for assessing RCTs to be used by authors, referees, and readers. Other journals[59] have published their policy on the statistical assessment of trials. With the possible exception of the *British Journal of Obstetrics and Gynaecology*,[6] these efforts have not had their intended impact.[60] We believe our proposal for structured reporting will improve the quality of reporting of RCTs. Our proposal may be unique in that it identifies which items should be included, why they should be included (empirical evidence, when available), and how they can be included (discussed hereinafter).

Structured reporting requires the reporting of what actually took place during the trial. We have provided a checklist of items that a group of trialists believe should be included in a structured report of a trial as well as why they need to be presented. Using this checklist, investigators will provide precise details about the design, conduct, and analysis of their trial. A principal advantage of such reporting is that all readers will have uniform and standardized information to review, unaffected by the writing nuances of authors and the policies of editors. This will give readers essential information about what happened during the trial, especially around issues affecting a trial's internal validity.

Some observers may argue that we have only included items, in particular those concerned with masking, that are more relevant to specific interventions, such as pharmacologic ones. Whether authors are reporting a triple-masked trial or a surgical trial in which it may be impractical or unethical to double-mask,[61] we believe it is important to report all the relevant items and justification if these were not carried out. In proposing structured reporting, our objective is not to pass judgment on the quality of the trial itself, but to improve on how it is reported to the reader.

Critics may view our proposal for structured reporting as rigid with a subsequent loss of the "creative process." On the contrary, our intent is not to discourage creativity in the research design process, but to encourage trial reports that provide readers with sufficient information. We believe there are several parts of a trial report, such as the introduction and discussion, where originality is needed. However, internal validity has little to do with the creative process and everything to do with providing accurate information to the reader about how the trial was performed. Describing the rationale for an intervention is just as important as describing how the participants were assigned to that intervention.

Where structured reporting will require inventiveness is in the style in which authors use it (ie, its format). One suggestion is for authors to report the methods and results section of their trial using the 24 essential items discussed above in a style similar to that of more informative abstracts. Each content area would be a separate heading with subheadings for each item within the content area. The number of items and subsequent text may vary slightly from report to report, especially for unmasked trials. What matters most when reporting trials using this approach is maintaining the structured report format.

Structured reporting, at least during the initial phase of its introduction, may increase the size of the trial report. It is also possible that structured reporting, with its focus on reporting issues relating to a trial's internal validity, may decrease the length of reports because other information, such as stating whether a trial was double-masked, can be removed as authors become more proficient in writing a structured report. Similar experiences were noted when more informative abstracts were first introduced.[62]

Our discussions indicated support for 24 recommendations and several areas where there is little or no evidence regarding the effects certain items have on estimates of intervention effects. There are important reasons to continue pursuing such methodologic investigations. The results of these studies will help to clarify which items should be added or deleted from a structured report and may influence which items should be included in any instrument developed to assess trial quality.

Although this report specifically addresses issues related to internal validity, several other items have been recommended to improve the reporting of RCTs. Laupacis and colleagues[63] have provided reasons for reporting results as absolute risk differences rather than relative risk reductions. Rochon et al[64] have recommended that age information be better reported, and Baar and Tannock[65] have encouraged the assessment and reporting of costs of therapy.

If structured reporting is to be successful, journal editors need to be involved to help refine the items to be included and to help with implementation and evaluation. This article documents the rationale for structured reporting of RCTs and suggests items to be included. We invite journal editors to review this article, join us in making it more suitable for "instructions to authors," and implement structured reporting of RCTs in their journals. We will continue to carry out discussions and empirical studies of the relations between trial reporting, trial quality, and trial results.

The Standards of Reporting Trials Group consisted of the following: Erik Andrew, PhD, Nycomed Imaging, Oslo, Norway; Aslam Anis, PhD, Clinical Epidemiology Unit, Ottawa (Ontario) Civic Hospital; Tom Chalmers, MD, New England Medical Center for Health Services Research and Study Design, Boston, Mass; Mildred Cho, PhD, Institute for Health Policy Studies, University of California, San Francisco; Mike Clarke, PhD, Clinical Trial Service Unit, Radcliffe Infirmary, Oxford, England; David Felson, MD (member of writing committee), Boston (Mass) University School of Medicine; Peter Gøtzsche, MD (member of writing committee), The Nordic Cochrane Centre, Research and Development Secretariat, Copenhagen, Denmark; Richard Greene, MD, PhD, Agency for Health Care Policy and Research, Rockville, Md; Alejandro Jadad, MD, Oxford Regional Pain Relief Unit, Headington, Oxford, England; Wayne Jonas, MD, Medical Research Fellowship, Walter Reed Army Institute of Research, Washington, DC; Terry Klassen, MD, Division of Emergency Medicine, Children's Hospital of Eastern Ontario, Ottawa; Paul Knipschild, MD, Ryksuniversiteit Limburg, Vakgoep Epidemiologie, Limburg, the Netherlands; Andreas Laupacis, MD, MSc (member of writing committee), Clinical Epidemiology Unit, Ottawa (Ontario) Civic Hospital; Curtis L. Meinert, PhD, Department of Epidemiology, The Johns Hopkins University, Baltimore, Md; David Moher, MSc (chair and member of writing committee), Clinical Epidemiology Unit, Ottawa (Ontario); Civic Hospital; Graham Nichol, MD, Clinical Epidemiology Unit, Brigham and Women's Hospital, Boston, Mass; Andy Oxman, MD, Department of Health Services, National Institute of Public Health, Oslo, Norway; Marie-France Penman, BScN, Multiple Sclerosis Clinic, Ottawa (Ontario) General Hospital; Stuart Pocock, PhD (member of writing committee), Medical Statistics Unit, London School of Hygiene and Tropical Medicine, University of London, (England); Joan Reisch, PhD,

Department of Pediatrics and Medical Computer Science, University of Texas Health Science Center, Dallas; David Sackett, MD (member of writing committee), Nuffield Department of Clinical Medicine, University of Oxford (England); Kenneth Schulz, PhD, MBA, Division of STD/HIV Prevention, National Center for Prevention Services, Centers for Disease Control and Prevention, Atlanta, Ga; Judy Snider, Department of Epidemiology and Community Medicine, University of Ottawa (Ontario); Peter Tugwell, MD (member of writing committee), Department of Medicine, University of Ottawa (Ontario); Jon Tyson, MD, Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas; France Varin, PhD, Faculty of Pharmacy, University of Montreal (Quebec); Wikke Walop, PhD, Biometrics Division, Drugs Directorie, Health Canada, Ottawa, Ontario; Sharon Walsh, MD, Department of Medicine, Ottawa (Ontario) Civic Hospital; and George Wells, PhD (member of writing committee), Clinical Epidemiology Unit, Ottawa (Ontario) Civic Hospital.

We would like to thank Du Pont Pharma, the National Health Research and Development Program, Health Canada, the Medical Research Council of Canada, Ortho-McNeil, and Rhône-Poulenc-Rorer for their financial support in convening the workshop. We would also like to thank Francis Rolleston for his encouragement, Carol Reichert for reviewing an earlier draft of the manuscript, and Iain Chalmers for his support in this process and his thoughtful comments on the manuscript.

References

1. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest.* 1992;102: 305S-311S.
2. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. *Am J Med.* 1982;72:233-240.
3. Mahon WA, Daniel EE. A method for the assessment of reports of drug trials. *Can Med Assoc J.* 1964;90:565-569.
4. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. *N Engl J Med.* 1987;317:426-432.
5. Altman DG, Doré CJ. Randomization and baseline comparisons in clinical trials. *Lancet.* 1990;335: 149-153.
6. 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;272:125-128.
7. Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis. *Control Clin Trials.* 1989;10:31-56. Correction. 1989;10:356.
8. Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. *Ann Intern Med.* 1987;106:598-604.
9. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med.* 1990;113:69-76.
10. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA.* 1994;272:122-124.
11. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials.* In press.
12. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Emperical evidence of bias: dimensions of methodologic quality associated with estimates of treatment effects in controlled trials. *JAMA.* In press.
13. Schulz KF. *Methodological Quality and Bias in Randomized Controlled Trials.* London, England: University of London; 1994. Thesis.
14. Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci.* 1993;703:156-165.
15. Gøtzsche PC. Bias in double-blind trials. *Dan Med Bull.* 1990;37:329-336.
16. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires.* New York, NY: Oxford University Press; 1987.
17. Steiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* Oxford, England: Oxford University Press; 1989.
18. Donner A, Brown KS, Brasher P. A methodologic review of nontherapeutic intervention trials employing cluster randomization, 1979-1989. *Int J Epidemiol.* 1990;19:795-800.
19. Donner A. Statistical methodology for paired cluster designs. *Am J Epidemiol.* 1987;126:972-979.
20. Altman DG. Randomization: essential for reducing bias. *BMJ.* 1991;302:1481-1482.
21. Hill AB. The clinical trial. *N Engl J Med.* 1952; 247:113-119.
22. Chalmers TC, Celano P, Sachs HS, Smith H. Bias in treatment assignment in controlled clinical trials. *N Engl J Med.* 1983;309:1358-1361.
23. Last JM, ed. *A Dictionary of Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 1988:38.
24. Altman DG. Comparability of randomized groups. *Statistician.* 1985;34:125-136.
25. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy, I: medical. *Stat Med.* 1989;8:441-454.
26. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold: a prophylactic and therapeutic trial. *JAMA.* 1975;231:1038-1042.
27. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med.* 1979;301:1410-1412.
28. Sackett DL. Bias in analytic research. *J Chronic Dis.* 1979;32:51-63.
29. The Canadian Cooperative Study Group. The Canadian trial of aspirin and sulfinpyrazone in threatened stroke. *N Engl J Med.* 1978;299:53-59.
30. Moses LE, Emerson JD, Hosseini H. Analyzing data from ordered categories. *N Engl J Med.* 1984; 311:442-448
31. Fields WS, Maslenikov V, Mayer 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.
32. Pocock SJ. *Clinical Trials: A Practical Approach.* Chichester, England: John Wiley & Sons Inc; 1983:182-186.
33. Laupacis A. The validity of survivorship analysis in total joint arthroplasty. *J Bone Joint Surg Am.* 1989;71:1111-1112.
34. Godfrey K. Comparing the means of several groups. In: Bailer JC III, Mosteller F, eds. *Medical Uses of Statistics.* 2nd ed. Boston, Mass: NEJM Books; 1992:233-257.
35. Makuch R, Johnson M. Issues in planning and interpreting active control equivalence studies. *J Clin Epidemiol.* 1989;42:503-511.
36. The Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *N Engl J Med.* 1980;303:1038-1041.
37. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med.* 1991;10:1595-1605.
38. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. In: Bailer JC III, Mosteller F, eds. *Medical Uses of Statistics.* 2nd ed. Boston, Mass: NEJM Books; 1992: 333-347.
39. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial: survey of 71 'negative' trials. *N Engl J Med.* 1978;299:690-694.
40. Laupacis A, Connolly SJ, Gent M, Roberts RS, Cairns JA, Joyner C, for the CAFA Study Group. How should results from completed studies influence ongoing clinical trials? the CAFA Study experience. *Ann Intern Med.* 1991;115:818-822.
41. Pocock SJ. When to stop a clinical trial. *BMJ.* 1992;305:235-240.
42. Evans M, Pollack AV. Trials on trial: a review of trials of antibiotic prophylaxis. *Arch Surg.* 1984; 119:109-113.
43. Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the *British Medical Journal. JAMA.* 1990;263:1355-1357.
44. McKinney WP, Young MJ, Hartz A, Bi-Fong Lee M. The inexact use of Fisher's Exact Test in six major medical journals. *JAMA.* 1989;261:3430-3433.
45. Gardner MJ, Altman DG. Estimation rather than hypothesis testing: confidence intervals rather than P values. In: Gardner MJ, Altman DG, eds. *Statistics With Confidence: Confidence Intervals and Statistical Guidelines.* London, England: BMJ; 1989:6-19.
46. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in participants receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991;324: 781-788.
47. Braitman LE. Confidence intervals assess both clinical significance and statistical significance. *Ann Intern Med.* 1991;114:515-517.
48. Amery A, Roeber G, Vermeulen HJ, Verstraete M. Single-blind randomized multicentre trial comparing heparin and streptokinase treatment in recent myocardial infarction. *Acta Med Scand.* 1969; 505:1-35.
49. Steering Committee of the Physicians' Health Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med.* 1989;321:129-135.
50. Bauer P. Multiple testing in clinical trials. *Stat Med.* 1991;10:871-890.
51. Bulpitt CJ. Subgroup analysis. *Lancet.* 1988;2: 31-34.
52. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med.* 1992;116:78-84.
53. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA.* 1991;266:93-98.
54. Simon R, Wittes RE. Methodologic guidelines for reports of clinical trials. *Can Treat Rep.* 1985;69:1-3.
55. Grant A. Reporting controlled trials. *Br J Obstet Gynaecol.* 1989;96:397-400.
56. Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *J Clin Oncol.* 1983;1:164-169.
57. Mosteller F, Gilbert JP, McPeek B. Reporting standards and research strategies for controlled trials. *Control Clin Trials.* 1980;1:37-58.
58. Gardner MJ, Machin D, Campbell MJ. Use of checklists in assessing the statistical content of medical studies. *BMJ.* 1986;292:810-812.
59. Gore SM, Jones G, Thompson SG. The *Lancet*'s statistical review process: areas for improvement by authors. *Lancet.* 1992;340:100-102.
60. Altman DG. Statistics in medical journals: developments in the 1980s. *Stat Med.* 1991;10:1897-1913.
61. Turner JA, Deyo RA, Loeser JD, Von Korff M, Fordyce WE. The importance of placebo effects in pain treatment and research. *JAMA.* 1994;271:1609-1614.
62. Comans ML, Overbeke AJ. The structured summary: a tool for reader and author. *Ned Tijdschr Geneeskd.* 1990;134:2338-2343.
63. Laupacis A, Naylor CD, Sackett DL. How should the results of clinical trials be presented to clinicians? *ACP Journal Club.* 1992;A12-A14.
64. Rochon PA, Fortin PR, Dear KBG, Minaker KL, Chalmers TC. Reporting of age data in clinical trials of arthritis: deficiencies and solutions. *Arch Intern Med.* 1993;153:243-248.
65. Baar J, Tannock I. Analyzing the same data in two ways: a demonstration model to illustrate the reporting and misreporting of clinical trials. *J Clin Oncol.* 1989;7:969-978.