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. 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 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 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 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 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 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. 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. More recently, there has been a call to extend more informative abstracts to include structured reporting of the text of each RCT.

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, 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 reviewed 250 reports of RCTs and found the odds ratios in the unclearly concealed trials are on average 30% (95% confidence interval, 21% to 38%) lower than in the adequately concealed trials.
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 82 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.

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.25 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.26 One of the principal advantages of randomization, noted by Hill,22 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 known confounders, permitting unbiased comparisons between intervention groups.26

3. Describe the Method Used to Conceal the Intervention Assignment Schedule From Participants and Clinicians Until Recruitment Was Complete and Irreversible.—To avoid selection bias in participant assignment to intervention, concealment is essential and should be feasible in all trials. Chalmers and colleagues22 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 un-concealed (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

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Unable to Determine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. State the unit of assignment.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. State the method used to generate the intervention assignment schedule.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Describe the method used to conceal the intervention assignment schedule from participants and clinicians until recruitment was complete and irreversible.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Describe the method(s) used to separate the generator and executor of the assignment.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Describe an auditable process of executing the assignment method.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Identify and compare the distributions of important prognostic characteristics and demographics at baseline.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. State the method of masking.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. State how frequently care providers were aware of the intervention allocation, by intervention group.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. State how frequently participants were aware of the intervention allocation, by intervention group.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. State whether (and how) outcome assessors were aware of the intervention allocation, by intervention group.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. State whether the investigator was unaware of trends in the study at the time of participant assignment.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. State whether masking was successfully achieved for the trial.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. State whether the data analyst was aware of intervention allocation.*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. State whether individual participant data were entered into the trial database without awareness of intervention allocation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. State whether the data analyst was masked to intervention allocation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Describe fully the numbers and flow of participants, by intervention group, throughout the trial.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. State clearly the average duration of the trial, by intervention group, and the start and closure dates for the trial.†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Report the reason for dropout clearly, by intervention group.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Describe the actual timing of measurements, by intervention group.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. State the predefined primary outcome(s) and analyses only.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Describe clearly whether the primary analysis has used the intention-to-treat principle.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. State the intended sample size and its justification.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. State and explain why the trial is being reported now.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Describe and/or compare trial dropouts and completers.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. State or reference the reliability, validity, and standardization of the primary outcome.‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. Define what constituted adverse events and how they were monitored by intervention group.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. State the appropriate analytical techniques applied to the primary outcome measure(s).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. Present appropriate measures of variability (eg, confidence intervals for primary outcome measures).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. Present sufficient simple (unadjusted) summary data on primary outcome measures and important side effects so that the reader can reproduce the results.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. State the actual probability value and the nature of the significance test.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. Present appropriate interpretations (eg, NS, no effect; P&lt;0.05, proof)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. Present the appropriate emphasis in displaying and interpreting the statistical analysis, in particular controlling for unplanned comparisons.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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


Structured Reporting of RCTs—The Standards of Reporting Trial's Group

1927
randomization,\(^\text{17}\) would result in the control group receiving 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.\(^\text{18,19}\) 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.\(^\text{18}\) 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 merely represent their suspicion about efficacy. For example, before the Canadian aspirin trial,\(^\text{20}\) 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.\(^\text{21}\) Gøtzsche\(^\text{1}\) 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.

**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.\(^\text{17}\) 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.\(^\text{18,19}\) Investigators should also report on how they assessed whether caregivers knew the intervention group to which the participants were being assigned.

**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.\(^\text{18}\) Investigators should also report on how they assessed whether participants knew their assignment group.

**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 merely represent their suspicion about efficacy. For example, before the Canadian aspirin trial,\(^\text{20}\) 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.

**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.\(^\text{21}\) Gøtzsche\(^\text{1}\) 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.

**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.\(^\text{17}\) 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.\(^\text{18,19}\) Investigators should also report on how they assessed whether caregivers knew the intervention group to which the participants were being assigned.

**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.\(^\text{18}\) Investigators should also report on how they assessed whether participants knew their assignment group.

**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 merely represent their suspicion about efficacy. For example, before the Canadian aspirin trial,\(^\text{20}\) 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.

**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.\(^\text{21}\) Gøtzsche\(^\text{1}\) 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.
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. 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. Analysis based on the intervention participants actually received, rather than the intervention to which they were initially assigned (efficacy analysis), may produce invalid results. 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. 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.

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.

Results Section

27. State the Appropriate Analytic Techniques Applied to the Primary Outcome Measure(s).—Evans and Follois 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%). 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. 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, particularly at the beginning.

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.

29. Present 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, \( P < 0.05 \)). Although the probability value \( P = 0.06 \) is not statistically significant at the 5% level and \( P = 0.04 \) 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, \( P = 0.02, z = 2.33 \)). 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, 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.

31. Present Appropriate Interpretations (eg, NS, No Effect; \( P < 0.05 \), 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. 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.

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.

Subgroup analysis is commonly conducted and inappropriately interpreted in many RCTs. Such analyses are more credible 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. Oxman and Guyatt and Yusuf and colleagues 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 to facilitate this process, describing what needs to be included when reporting a trial. In addition, some journals have published checklists of items for assessing RCTs to be used by authors, referees, and readers. Other journals have published their policy on the statistical assessment of trials. With the possible exception of the *British Journal of Obstetrics and Gynaecology,* these efforts have not had their intended impact. 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 pharmacoeconomic ones. Whether authors are reporting a triple-masked trial or a surgical trial in which it may be impractical or unethical to double-mask, 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 increase 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 starting whether the 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.

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 have provided reasons for reporting results and the risk differences rather than relative risk reductions. Rochon et al have recommended that age information be better reported, and Baar and Tannoel 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.