CONSORT guidelines applied to an exemplar paper

Introduction

The accompanying editorial (Newcombe 2004) describes the rationale for the CONSORT guidelines and announces their adoption as *International Endodontic Journal* policy. The editorial board agreed it would also be helpful to present an example of CONSORT applied to a paper already published in the *International Endodontic Journal* and to show how good research could be further strengthened. The paper by Weiger *et al.* (2000) was suggested as a suitable exemplar, and the authors consented to their paper being used in this way.

The purpose of the present article is to illustrate the application of the CONSORT guidelines to a published paper by attempting to reconstruct the CONSORT checklist and flowchart for it. It must be emphasized that CONSORT was designed for prospective use, as described below, and not to enable a reviewer to assess the quality of a paper retrospectively – no reviewer will have the full information available to the authors. Nevertheless, we hope that readers will benefit from this attempt to show what is required to make one of the better studies into one that would be regarded as yielding the highest level of evidence.

The two key elements of CONSORT are a checklist and a flowchart. The flowchart is intended to form part of the published paper. The checklist is not, but is intended to be completed by the authors and submitted with the paper. It is a framework by which authors can convey, primarily to referees, the information that they need to check has been presented adequately, in order to judge whether the study is satisfactory. The flowchart shows what happens to all subjects who are considered for recruitment to the study. Some are excluded or do not consent to commence the study. Those who do are duly randomized. There are several ways in which these may not reach the endpoint and yield outcome information, and the numbers concerned should be shown for each treatment group separately.

Checklist

The reconstructed checklist for the exemplar paper is below as Table 1. The first four columns are copied from the CONSORT website. For clarity, the item 'Reported on page #' has been modified to 'Reported where?'. Pages and columns, tables and figures of the printed paper are identified as appropriate. The final column, 'Comments', gives my comments on whether this item has been handled in accordance with CONSORT – it is important to bear in mind that in normal, prospective use this column would not exist, as the purpose is for the authors to demonstrate to referees that they have dealt adequately with all these items within the manuscript.

Flowchart

Figure 1 is an attempt to reconstruct the flowchart for the exemplar study. Note that not all cells can be completed using the information given in the paper. This is for two reasons. Information is lacking on how many subjects or teeth were considered for recruitment but then excluded, and on what grounds.

Also, the paper refers to six patients who did not yield outcome data, five who failed to attend recall appointments or refused further examination and one death. We are not told which groups these patients were from, nor how many patients were actually randomized to each group (which would be implied by this information). It would be of some concern if all five refusals were in the two-visit group, as is quite plausible from the disparity between the numbers of completers on the two regimes, 36 and 31. If this is so, arguably these refusals should rather be considered as 'votes against' a two-visit treatment plan. This illustrates the importance of giving full information on dropouts, usually in flowchart form.

Discussion

The major weakness of the study overall is that it is small for its purpose and no rationale is given for the sample size used. The reviewer fully realizes that this study was a lot of work; nevertheless, it should be regarded as a very small one to give definitive information on the issue. This is best seen from the fact that the confidence intervals are very wide in Table 6. They would be wider still for the comparison of treatments, which should be (but is not) directly readable from Table 7. The study really gives very limited evidence on whether the choice between one- and two-stage regimes is important.

While retrospective power calculations to justify the sample size used in a study should not be regarded as

Table 1 Completed CONSORT checklist for exemplar study, with comments

Paper section and topic	Item	Description	Reported where?	Comments
Title and abstract	1	How participants were allocated to interventions (e.g. 'random allocation', 'randomized' or 'randomly assigned').	Page 221 column b	No mention of 'randomization', in title, abstract or body text. Minimization, used here, is an acceptable way to achieve this, but this paper risks not being picked up in a systematic review literature search designed to identify randomized trials for meta-analysis.
Introduction				
Background	2	Scientific background and explanation of rationale.	Page 219, 220a	OK.
Methods				
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	Page 220a & b	Eligibility criteria clear. Settings and locations where data collected not clear.
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	Page 220b, 221a	OK.
Objectives	5	Specific objectives and hypotheses.	Page 220a	Objective specific but not expressed as hypothesis.
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).	Page 221a & b	Clearly defined outcomes. Healing primary, though implicit rather than explicit. Both dentists read each radiograph.
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	–	No sample size rationale reported. Interim analyses, stopping rules evidently not used – little relevance in long-term follow-up study
Randomization – Sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (e.g. blocking, stratification).	Page 221b	Minimization reasonably explicit, though preferable to specify algorithm or software used. Restriction – none mentioned.
Randomization – 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.	–	No mention of allocation concealment.
Randomization – Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	Page 220b	By implication, RW and RR enrolled participants. No mention of who generated allocations, but less relevant as interactive minimization algorithm tends to work as a 'black box'. Who assigned – not mentioned.
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	Page 221a	Patient and operator blinding not achievable here. Assessor blinding mentioned for reading radiographs, but clinical assessments may not have been performed blind.

Table 1 continued

Paper section and topic	Item	Description	Reported where?	Comments
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses, such as subgroup analyses and adjusted analyses.	Page 221b, 222a	OK
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 analysed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	Page 220b	Clear how 6 patients recruited did not yield outcome data. Not clear how they distribute between the two groups.
Recruitment	14	Dates defining the periods of recruitment and follow-up.	Page 220a, 221b	OK
Baseline data	15	Baseline demographic and clinical characteristics of each group.	Tables 1–5, Fig. 1	Clinical characteristics described OK, but nothing on demographics.
Numbers analysed	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. 10/20, not 50%).	Page 222a & b	36 vs. 31, implicitly in all analyses. No protocol deviations necessitating heeding intention to treat principle. Absolute numbers are given.
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).	Tables 6 and 7	Table 6 OK. In Table 7, effect size and confidence interval only given when significant. This is poor practice in any case, and here results in suppressing the effect size information of primary interest.
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.	Table 8	Best and worst case scenario analyses in Table 8 (often described as sensitivity analyses) are a very positive feature. Impossible to judge whether any other analyses were performed/contemplated and not reported, but nothing obvious springs to mind.
Adverse events	19	All important adverse events or side-effects in each intervention group.	–	None reported.
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.	Pages 223b–225	Reasonable but ... see Discussion section below.
Generalizability	21	Generalizability (external validity) of the trial findings.		
Overall evidence	22	General interpretation of the results in the context of current evidence.		

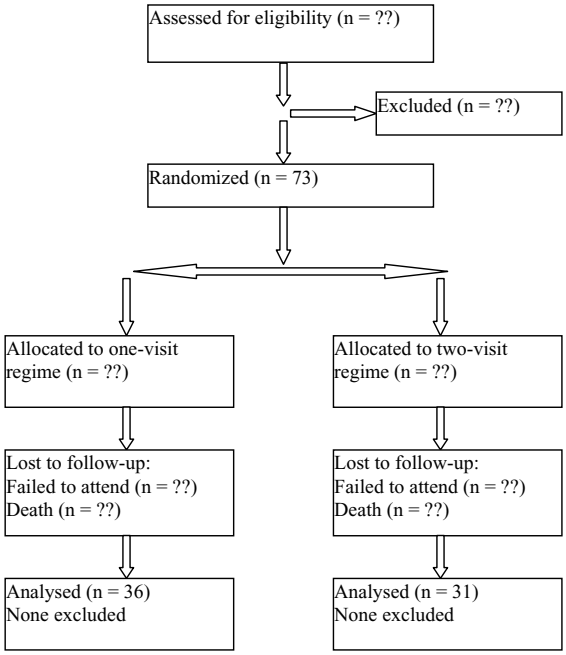

Figure 1 Reconstructed CONSORT flowchart for exemplar study.

acceptable practice, it is instructive to use the information given in this paper to assess what sample size would be appropriate in a study of this kind. The largest difference between groups was found at 3 years, with success rates 79 and 70% in the two groups.[1] To detect a difference between population proportions of 79 and 70% with 80% power at the conventional two-sided 5% significance level would require 389 subjects per group (nQuery Advisor 4.0). This is more than 10 times the number of subjects used in this study. Given that in the present study it took two clinicians 5 years to recruit 73 patients, it is clear that a considerably larger number of clinicians would be required to achieve an appropriate sample size. The resulting study would require very careful checks on consistency of operation of eligibility criteria, performance of surgical procedures and assessment. It is expected that substantial resources would be required to make it work.

The sample size issue points to one possible drawback of the CONSORT guidelines. Here, as in very many studies, the issue of small sample size and low power encapsulated in CONSORT item 7 is all-important, but the attention given to other issues could detract emphasis from it. Indeed, not all CONSORT items are as relevant to the study considered here as they would be to, say, a drug trial – nevertheless, all issues are worth giving thought to, even if just to dismiss them as not applicable.

Also, CONSORT does not cover all issues of relevance to all studies. Here, it is a very positive feature that just one tooth per patient is studied – it is well established that studies that disregard the non-independence of teeth, eyes, etc. in the same subject are prone to yield misleading results. This issue is highly relevant to all clinical studies in endodontics.

Another issue, of ethical rather than scientific importance in all studies on patients, is informed consent, which is not mentioned here. In an experimental study, it is necessary to obtain consent, not only to the treatment one intends to carry out just as per normal clinical practice but also to inclusion in a randomized trial, in which the treatment to be used is determined by other criteria in addition to the clinician's professional judgement.

Robert G. Newcombe
Reader in Medical Statistics, UWCM, Cardiff, UK
Statistical Advisor, IEJ Editorial Board

References

Newcombe RG (2004) Reporting of clinical trials in the *IEJ* – the CONSORT guidelines. *International Endodontic Journal* **37**, 1–2.

Weiger R, Rosendahl R, Löst C (2000) Influence of calcium hydroxide intracanal dressings on the prognosis of teeth with endodontically induced periapical lesions. *International Endodontic Journal* **33**, 219–26.

nQuery Advisor 4.0. Cork, Ireland: Statistical Solutions.

http://www.consort-statement.org/revisedstatement.htm (accessed 23/9/2003).

[1] One could ask, is a 9% difference large enough to be regarded as important? It would strengthen a study of this kind to appraise cost-effectiveness of one approach relative to the other.