The Journal of Laryngology and Otology
May 1998, Vol. **112**, pp. 460–463

A qualitative assessment of randomized controlled trials in otolaryngology

K. W. Ah-See, M.D., F.R.C.S., N. C. Molony, B.Sc., F.R.C.S.

Abstract

In 1996 the CONSORT statement made recommendations on the strict reporting of randomized controlled trials (RCT). This will facilitate the future assessment of such trials and will highlight those trials that have been performed suboptimally and whose results may be biased.

We have devised a scoring system, based on CONSORT, to assess RCT quality and by reading each original paper in full we have now assessed the quality of trials published from 1966 to 1995.

The mean score for trials identified was 7.3 out of a maximum 12 points. No one journal was significantly better than the others. Trials in rhinology are reported better than head and neck oncology trials (mean scores 7.6 and 6.5 respectively). The past 30 years has not seen an improvement in the quality of the trials.

The reporting of RCTs in the ENT literature is poor. CONSORT guidelines now exist and trialists are encouraged to adopt them when conducting future clinical trials.

Key words: Clinical trials; Reference standards; Otolaryngology

Introduction

If evidence-based medicine (EBM) is to be successfully implemented within the specialty of ENT then, by definition, this demands the marriage of two crucial aspects; firstly, knowledge gained from clinical expertise and experience and secondly, a robust body of external scientific evidence. The former arises out of good training and years in the job. The latter, however, is perhaps harder to achieve and currently centres around the gold standard RCT.

We have recently reported an encouraging trend in the absolute numbers of RCTs that have appeared in the ENT literature over a 30-year period (Ah-See *et al.*, 1997). In 1996 however the CONSORT statement made it clear on both sides of the Atlantic that strict guidelines were to be followed in the future reporting of RCTs (Altman, 1996; Begg *et al.*, 1996). The reasons for this are simple. If the results of an RCT were to be valid then trials should be well executed, unbiased and the reader should be told explicitly what was done rather than have to infer what was probably done (Altman, 1996). Only valid results from good quality RCTs should be used to devise evidence-based clinical guidelines.

We have, therefore, qualitatively assessed and scored each of the previously reported 295 RCTs that were highlighted in a Medline search of the ENT literature between 1966 and 1995. Our aim was to identify how well these trials have been performed on the basis of what is reported in the full article compared to what is now being demanded (Altman, 1996). We anticipated being able to highlight areas, within the reporting of these trials, that were lacking and thus devise recommendations on the future conduct and reporting of RCTs within our specialty.

Materials and methods

A 30-year Medline search was performed from 1966 to 1995. Our search strategy was based on identifying the textwords, randomized, control, controlled, trial, study, prospective, RCT, and clinical trial.

The search was executed on 10 prominent ENT journals (Table I) any journal supplements were also included in the search. Our choice of journals was based on high circulation numbers and citation indices (*Science Citation Index*, 1992). Not all journals have been in existence since 1966, for

TABLE I
ENT JOURNALS SEARCHED

- *Acta Otolaryngologica*
- *Annals of Otology, Rhinology and Laryngology*
- *Archives of Otolaryngology – Head and Neck Surgery*
- *Audiology*
- *British Journal of Audiology*
- *Clinical Otolaryngology*
- *Journal of Laryngology and Otology*
- *Laryngoscope*
- *Otolaryngology – Head and Neck Surgery*
- *Rhinology*

From the Department of Otolaryngology, The Royal Infirmary of Edinburgh, Edinburgh, UK.
Accepted for publication: 16 March 1998.

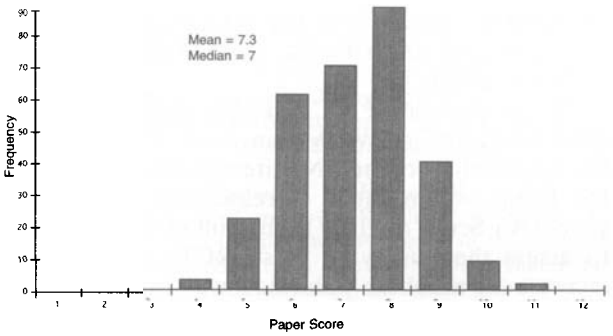

FIG. 1

Distribution of scores for the RCTs identified.

example Clinical Otolaryngology was first published in 1976. All abstracts highlighted from the search were read to identify the article as an RCT. The full papers were in turn obtained and read in full to assess the trial quality. The quality assessment was performed using a scoring system derived from the recommendations laid down by the CONSORT statement (Begg *et al.*, 1996) and from observations previously made on the ENT literature (Rosenfeld and Rockette, 1991). The scoring proforma (Appendix) had a maximum score of 12 points covering the salient features of a study. It is intentionally a condensed form of the CONSORT recommendations as we wanted a succinct yet convenient means to assess the near 300 full articles that we read. The salient features, however, of the CONSORT statement were maintained.

All articles were read by one or other author. Fifty computer-generated random numbers were used to choose papers to be read by both authors to assess inter-observer concordance in the quality assessment.

Results

The initial Medline search highlighted 370 articles in the 10 journals from 1966–95. Further analysis of the abstracts, however, revealed 75 articles that clearly were not trials but merely included the textwords in the abstract such as: ' . . . we recommend a randomized controlled trial . . . '.

A total of 295 original articles of an RCT design were analysed. Each paper was read in full and scored using the proforma (Appendix). The mean score for all trials was 7.3 out of 12 (range 4 to 11). No article obtained the full 12 point score (Figure 1).

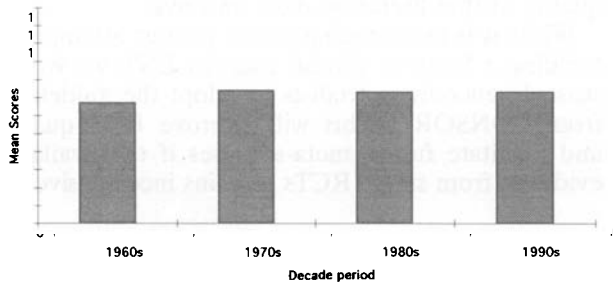

FIG. 2

Mean scores for the articles identified per decade.

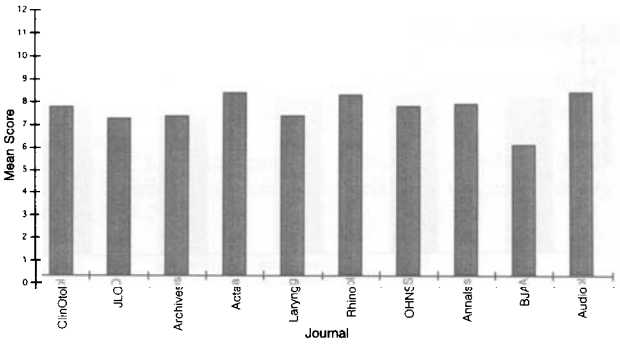

FIG. 3

Mean scores for the articles identified per journal searched.
ClinOtol: *Clinical Otolaryngology*
JLO: *The Journal of Laryngology and Otology*
Archives: *Archives of Otolaryngology – Head and Neck Surgery*
Acta: *Acta Otolaryngologica*
Laryng: *Laryngoscope*
Rhinol: *Rhinology*
OHNS: *Otolaryngology – Head and Neck Surgery*
Annals: *Annals of Otology, Rhinology and Laryngology*
BJA: *British Journal of Audiology*
Audiol: *Audiology*

A breakdown of the mean scores for the articles over consecutive decades showed no significant change after the 1970s. Only three papers were found for the 1960s and their mean score was 6.7. Thereafter the scores from the 1970s through the 1980s to the 1990s were, 7.4, 7.3 and 7.3 respectively (Figure 2).

The mean score for the individual journals are shown in Figure 3 (range 5.7 to 8.0). Few trials were found in the journals *Audiology* (two trials) and *British Journal of Audiology* (three trials) hence their mean scores of 8 and 5.7 respectively reflect this (Figure 4).

The majority of trials, in the journals searched, were within otology and rhinology (Ah-See *et al.*, 1997) presumably for the following reasons: ear and nose complaints constitute the majority of general ENT referrals; medical treatments play a large role in conditions affecting the ear and nose; the patient can act as his/her own control thus facilitating the RCT design.

The average scores for the trials performed within each broad category of ENT are shown in Figure 5. Head and neck oncology trials had lower scores (mean 6.5, range 5–9) than rhinology trials (mean 7.6, 5–11).

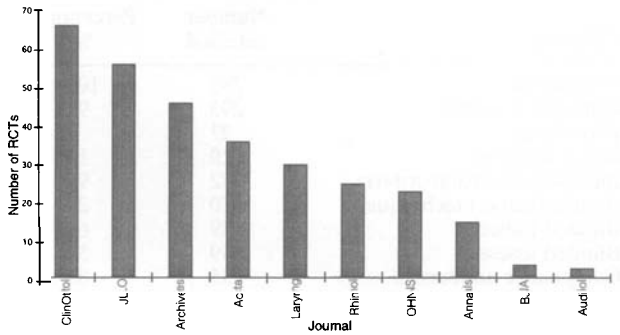

FIG. 4

Number of RCTs published per journal, 1966–1995.

https://doi.org/10.1017/S0022215100140770 Published online by Cambridge University Press

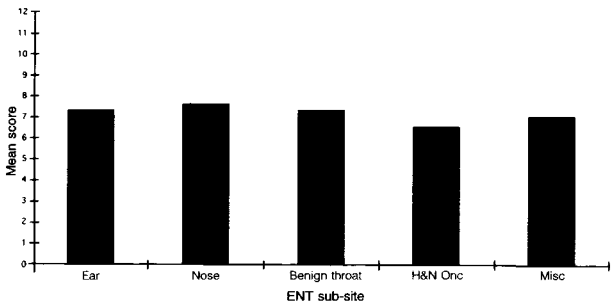

FIG. 5

Mean scores for the articles within ENT sub-sites. H & N Onc = Head and Neck Oncology; Misc = Miscellaneous.

We acknowledge that the journals searched did not specifically include any oncology journals and therefore we anticipate a number of head and neck oncology trials will have been missed. This review however was not intended as an exhaustive search more as a representative sample.

Table II indicates the number of trials satisfying the scoring criteria. The most consistent deficiencies were in piloting of trials, power analysis, randomization techniques and reporting of confidence intervals.

Inter-observer concordance was high with both authors scoring 47 of 50 random articles the same (94 per cent concordance). The three papers where scoring was not concordant differed by only one scoring point and was usually due to one or other author overlooking the particular point despite it being mentioned in the text of the paper.

Discussion

The demand for evidence-based clinical practice continues to grow. It is clear from reading the medical literature that there is an increasing number of randomized controlled trials being published. This study-design remains the gold standard for sound external, scientific evidence (Chalmers *et al.*, 1992).

The CONSORT statement published in 1996 set out criteria that should be included in the reporting of an RCT to allow the reader to qualitatively assess the conduct of the trial and hence the results (Begg *et al.*, 1996). Prominent medical journals will in future reject clinical trials that do not conform to the standards laid down by the CONSORT statement (Altman, 1996).

TABLE II

THE NUMBER AND PERCENTAGE OF ARTICLES SATISFYING THE 12 SCORING CRITERIA

Criterion	Number satisfied	Percentage %
Hypothesis	295	100
Outcome measure	293	99
Pilot study	27	9
Power analysis	29	10
Inclusion/exclusion criteria	282	96
Randomization technique	70	24
Blinded patient	189	64
Blinded assessor	159	54
Compliance addressed/ Handling of withdrawals	255	86
Statistics appropriate	238	81
Confidence intervals	26	9
Conclusions valid	291	99

The proportion of RCT articles published in the ENT literature is low (Maran *et al.*, 1997). Within this small subset of the ENT literature however there has been a quantitative increase over the past 30 years (Ah-See *et al.*, 1997). The aim of this study was to assess the quality of these RCTs to see if the increase in numbers published was mirrored by a trend in trial quality. A Medline search can only ever be a sample of the whole literature nonetheless we feel it is sufficiently representative to allow comment.

Our assessment, using a scoring system derived from the CONSORT criteria, has highlighted a disappointing level of quality in the reporting of these trials. Solomon *et al.* found similar results when analysing the quantity and quality of trials in the general surgical literature (Solomon *et al.*, 1994).

In our study an overall mean score of 7.3 out of a possible 12 is unsatisfactory. This implies the omission of between four and five criteria in the RCTs reported in the ENT literature. The most commonly noted failings are in the areas of study piloting, power analysis, inclusion of confidence intervals in the results and description of the randomization technique. It is no longer sufficient to say: ' . . . patients were randomized in to groups . . . '. A clear description of the method of randomization is required, e.g. closed envelopes, hospital unit number, tossing a coin etc. The description and details of a power analysis, calculated prior to embarking on a study, is important if the reader is to assess whether the study sample size is sufficient to show a difference, if one truly exists. Likewise the inclusion of confidence intervals in the results allows an assessment of how valid and precise the reported results are.

Although there has been an increase in numbers of published trials this has not unfortunately been paralleled by improved quality. Trials nowadays are not being performed or reported any better than 20 years ago (Figure 2). The introduction of the CONSORT statement may now herald a change in this situation with future years witnessing improved reporting of RCTs.

If bodies such as the Cochrane collaboration (Chalmers *et al.*, 1992) or the Scottish Intercollegiate Guidelines Network (SIGN) (1995) are to proceed with systematic reviews of the ENT literature in an attempt to develop robust clinical guidelines then the quality of that literature must improve.

While it is encouraging to see greater attempts at building a body of clinical trials in ENT we would strongly encourage trialists to adopt the guidelines from CONSORT. This will improve RCT quality and facilitate future meta-analyses if the available evidence from single RCTs remains inconclusive.

Acknowledgements

We would like to thank Professor A. G. D. Maran for his advice in the preparation of this manuscript.

References

Ah-See, K. W., Molony, N. C., Maran, A G. D. (1997) Trends in randomized controlled trials in ENT: a 30 year review. *Journal of Laryngology and Otology* **111:** 611–613.

Altman, D. G. (1996) Better reporting of randomised controlled trials: the CONSORT statement. *British Medical Journal* **313:** 570–571.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D., Stroup, D. F. (1996) Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *Journal of the American Medical Association* **276:** 637–639.

Chalmers, I., Dickersin, K, Chalmers, T. C. (1992) Getting to grips with Archie Cochrane's agenda (editorial). *British Medical Journal* **305:** 786–788.

Journal Citation Reports (1992) Subject category listing, Otorhinolaryngology. In *Science Citation Index*, Institute for Scientific Information Inc., Philadelphia, USA, p 108.

Maran, A. G. D., Molony, N., Armstrong, M., Ah-See, K. W. (1997) Is there an evidence base for the practice of ENT surgery? *Clinical Otolaryngology* **22(2):** 152–157.

Rosenfeld, R. M., Rockette, H. E. (1991) Biostatistics in Otolaryngology journals. *Archives of Otolaryngology – Head and Neck Surgery* **117:** 1172–1176.

SIGN, Scottish Intercollegiate Guidelines Network (SIGN). (1995) The Royal College of Physicians, Queen Street, Edinburgh.

Solomon, M. J., Laxamana, A., Devore, L., McLeod, R. S. (1994) Randomized controlled trials in surgery. *Surgery* **115(6):** 707–712.

Address for correspondence:
Dr Kim Ah-See M.D., F.R.C.S.,
Department of Otolaryngology,
The Royal Infirmary of Edinburgh,
Edinburgh EH3 9YW.

Fax: 0131-536-6167

APPENDIX

Title:

Journal Reference:

				Y	N
1.	**Endpoints/Outcome Measures**	Defined?		☐	☐
	Hypothesis?			☐	☐
2a.	**Pilot** study? (or any mention of being piloted?)			☐	☐
b.	**Power** Calculation	Mentioned/Attempted?		☐	☐
c.	**Number*** in groups	Control		.....	
		Treatment		.....	
3.	**Inclusion/Exclusion** Criteria			☐	☐
4.	**Control** Group*	Placebo ☐	Is a no treatment group ethical?	☐	☐
		Another treatment ☐			
5a.	**Randomization** Technique described (in detail)?			☐	☐
		If YES, what was it?			
b.	**Blinding**?	Patients?		☐	☐
		Doctors/Assessor		☐	☐
6.	Is **Handling of Withdrawals** (inc. patient compliance) addressed?			☐	☐
7.	**Statistics**?	What test?			
		Appropriate?		☐	☐
		Confidence intervals?		☐	☐
8.	Are **Conclusions** appropriate for the results (i.e. valid interpretation)			☐	☐

(Y = 1; N = 0) **TOTAL** (of 12) .....

(* = non-scoring questions)

Quality Scoring Form derived from CONSORT recommendations.

https://doi.org/10.1017/S0022215100140770 Published online by Cambridge University Press