

An Evaluation of the Quality of Clinical Trials in Anesthesia

Hwee Leng Pua, M.B.B.S., M.Med.,* Jerrold Lerman, B.A.Sc., M.D., F.R.C.P.C., F.A.N.Z.C.A.,†
Mark W. Crawford, M.B.B.S., F.R.C.P.C.,‡ James G. Wright, M.D., F.R.C.S.C., M.P.H.§

Background: The authors evaluated the quality of clinical trials published in four anesthesia journals during the 20-yr period from 1981–2000.

Methods: Trials published in four major anesthesia journals during the periods 1981–1985, 1991–1995, and the first 6 months of 2000 were grouped according to journal and year. Using random number tables, four trials were selected from all of the eligible clinical trials in each journal in each year for the periods 1981–1985 and 1991–1995, and five trials were selected from all of the trials in each journal in the first 6 months of 2000. Methods and results sections from the 160 trials from 1981–1985 and 1991–1995 were randomly ordered and distributed to three of the authors for blinded review of the quality of the study design according to 10 predetermined criteria (weighted equally, maximum score of 10): informed consent and ethics approval, eligibility criteria, sample size calculation, random allocation, method of randomization, blind assessment of outcome, adverse outcomes, statistical analysis, type I error, and type II error. After these trials were evaluated, 20 trials from the first 6 months of 2000 were randomly ordered, distributed, and evaluated as described.

Results: The mean (\pm SD) analysis scores pooled for the four journals increased from 5.5 ± 1.4 in 1981–1985 to 7.0 ± 1.1 in 1991–1995 ($P < 0.00001$) and to 7.8 ± 1.5 in 2000. For 7 of the 10 criteria, the percentage of trials from the four journals that fulfilled the criteria increased significantly between 1981–1985 and 1991–1995. During the 20-yr period, the reporting of sample size calculation and method of randomization increased threefold to fourfold, whereas the frequency of type I statistical errors remained unchanged.

Conclusion: Although the quality of clinical trials in four major anesthesia journals has increased steadily during the past two decades, specific areas of trial methodology require further attention.

THE randomized clinical trial is the highest level of evidence available for evaluating new therapies.¹ However, serious deficiencies in study design and data anal-

ysis have been reported in reviews of clinical trials in medical and surgical journals. DerSimonian *et al.*² noted that approximately half of the criteria deemed to be essential qualities of good study design were either ambiguously reported or not reported at all in their survey of 67 clinical trials from four medical journals published between July 1979 and June 1980. Emerson *et al.*³ reported similar results in a survey of clinical trials published in general surgical journals between 1981 and 1982. These deficiencies in study design introduce an element of bias into the trial results, which may lead to exaggerated treatment effects.^{4,5} To address these deficiencies in study design and preclude misleading results, criteria for excellence in randomized clinical trials have been published.^{6–9} Despite these publications, the quality of clinical trials in virtually all contexts that have been assessed has not improved substantively.^{10–15} For example, Emerson *et al.*¹¹ observed that the quality of trials in drug development improved only 9% in each of the past three decades. With the ever-increasing dependency of new interventions in anesthesia on evidenced-based outcome trials, we hypothesized that the quality of trials in the anesthesia literature has improved in the past 20 yr. Accordingly, we evaluated the quality of the study design in prospective clinical trials that were published in four leading anesthesia journals during the 20-yr period from 1981–2000.

Methods

Each issue of four anesthesia journals, *Anesthesia and Analgesia*, *ANESTHESIOLOGY*, the *British Journal of Anaesthesia*, and the *Canadian Journal of Anaesthesia*, was manually searched for all prospective comparative clinical trials published between 1981 and 1985, between 1991 and 1995, and during the 6-month period from January to June 2000. Trials that involved animals and human volunteers, evaluation of equipment, comparative trials using case, cohort, or historical controls, and retrospective and observational studies were excluded. Of the trials from the periods 1981–1985 and 1991–1995 that satisfied the eligibility criteria, four were randomly selected from each journal for each of the 10 yr being considered using random number tables. A total of 160 clinical trials were selected for review by one of the authors who did not participate in the review process. The methods and results sections from each trial were selectively copied, blinded to the journal and year of publication, and then randomly ordered using random

This article is accompanied by an Editorial View. Please see:
Todd MT: Clinical research manuscripts in ANESTHESIOLOGY.
ANESTHESIOLOGY 2001; 95:1051–3.

* Fellow, † Siemens Chair in Paediatric Anaesthesia and Professor of Anaesthesia, University of Toronto, Toronto, Ontario, Canada. Submitted for publication April 10, 2001. Accepted for publication June 26, 2001. Dr. Wright is supported by the Canadian Institute for Health Research, Ottawa, Canada. Presented in part at the annual meeting of the Canadian Anaesthetists' Society, Calgary, Alberta, Canada, June 5, 1999.

Received from The Hospital for Sick Children and the Research Institute, University of Toronto, Toronto, Ontario, Canada. Submitted for publication April 10, 2001. Accepted for publication June 26, 2001. Dr. Wright is supported by the Canadian Institute for Health Research, Ottawa, Canada. Presented in part at the annual meeting of the Canadian Anaesthetists' Society, Calgary, Alberta, Canada, June 5, 1999.

Address reprint requests to Dr. Lerman: Department of Anaesthesia, Hospital for Sick Children, Toronto, M5G 1X8, Ontario, Canada. Address electronic mail to: lerman@anaes.sickkids.on.ca. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

number tables. The randomization codes were concealed until the review process was completed. Each trial was examined for 10 characteristics of quality related to the study design by three of the authors (J. L., M. W. C., and J. G. W.) independently.² The rating system modified from that of DerSimonian *et al.*² was used because it was most appropriate for the type of clinical trials conducted in anesthesia:

1. informed consent and ethics approval: statements reflecting institutional ethics committee approval and informed consent from the patient or guardian
2. eligibility criteria: inclusion and exclusion criteria were provided for patients in the trial
3. sample size calculation: a sample size calculation or specification for the size of detectable differences before the trial commenced was provided
4. random allocation: a statement or phrase that indicated that the patients were assigned to their treatments in a randomized manner
5. method of randomization: details about the mechanism used to generate random assignment; randomization by birth date, hospital registration numbers, and so forth were considered unacceptable
6. blind assessment of outcome: a statement that the observer who assessed the outcome was unaware of the treatment assignment
7. adverse outcomes: details of the presence or absence of side effects or complications, not related to the primary outcome variable
8. statistical analysis: appropriate statistical tests were listed
9. type I error: defined as more than 10 comparisons among the treatment groups without correction or consideration of multiple tests
10. type II error: for trials with no statistically significant differences between groups in which neither a sample size estimation nor a power analysis was performed

When the reviews for the first two periods were completed, five prospective comparative clinical trials were randomly selected from those published in the same four journals between January and June 2000 using random number tables. The methods and results sections of these trials were prepared as described and analyzed by the same three authors using the same scoring system.

For each trial, the 10 criteria were evaluated and reported as present, absent, or not applicable. Inadequate or ambiguous information was reported as absent. The 10 criteria were weighted equally. The first eight items were scored 1 if present and 0 if absent, whereas the last two items (type I and II errors) were scored 1 if absent and 0 if present. The scores for the 10 characteristics were summed to give a maximum score of 10 for each trial. Scoring discrepancies among the reviewers were

resolved by discussion and a consensus decision before the randomization code was revealed.

Statistical Analyses

For the trials published in 1981–1985 and 1991–1995, the frequency of reporting each characteristic (nominal data) was compared among journals and between years using the Fisher exact test. The total trial scores were compared among the journals and between the two review periods using two-factor analysis of variance with the Newman-Keuls test for multiple comparisons. Mean analysis scores for the four journals in the 1981–1985 and 1991–1995 periods were tested for normality using the Kolmogorov-Smirnov test. To account for the number of between-group comparisons (15 in total), the threshold for statistical significance was designated to be $P < 0.01$. The trials published in the first 6 months of 2000 were included for descriptive statistics; however, the numbers were too small to include them in the statistical analysis.

Results

All 180 trials were evaluated by the same three reviewers, and a consensus was reached on each criterion reviewed. For each of two periods, 1981–1985 and 1991–1995, twenty trials were reviewed from each of the four journals, giving a total of 160 trials. The frequencies of responses for each of the 10 criteria are summarized in table 1. Of the 10 criteria that were evaluated in each of the 160 trials (for a total of 1,600 evaluations), less than 1% were nonapplicable. For these few evaluations, the trial was omitted from the analysis of that criterion, and the number of journals in that period decreased accordingly. For the first 6 months of 2000, 20 trials, representing a mean sampling rate of 13% (range, 6–18%) of the published trials, were reviewed from the four journals. None of the criteria was deemed not applicable.

The mean analysis scores of the four journals in both the 1981–1985 and 1991–1995 periods were normally distributed. Within each period, the analysis scores for the four journals were similar and were therefore pooled (fig. 1). The pooled mean analysis scores increased significantly from 5.5 ± 1.4 in 1981–1985 to 7.0 ± 1.1 in 1991–1995 ($P < 0.00001$ compared with 1981–1985) and then to 7.8 ± 1.5 in 2000. The percentage of trials with scores of 9 or more out of 10 increased from 0% in 1981–1985 to 10% (3.5–16.5%, 95% confidence interval) in 1991–1995 ($P < 0.014$ compared with 1981–1985) and to 30% (11–49%, 95% confidence interval) in 2000.

Within the periods 1981–1985 and 1991–1995, the percentage of trials that fulfilled each criterion was similar among the four journals. Accordingly, we pooled their percentages into one value for the respective pe-

Table 1. Criteria for the Quality of the Study Design

Criterion	Journal				P values*
	A	B	C	D	
Consent	90 \Rightarrow 100 \Rightarrow 100	100 \Rightarrow 100 \Rightarrow 100	75 \Rightarrow 100 \Rightarrow 100	95 \Rightarrow 100 \Rightarrow 100	0.0065
Eligibility	65 \Rightarrow 95 \Rightarrow 100	85 \Rightarrow 100 \Rightarrow 100	70 \Rightarrow 90 \Rightarrow 100	90 \Rightarrow 100 \Rightarrow 100	0.0007
Sample size calculation	0 \Rightarrow 15 \Rightarrow 40	5 \Rightarrow 20 \Rightarrow 40	0 \Rightarrow 5 \Rightarrow 80	5 \Rightarrow 15 \Rightarrow 60	0.0174
Random allocation	90 \Rightarrow 100 \Rightarrow 80	58 \Rightarrow 95 \Rightarrow 100	90 \Rightarrow 100 \Rightarrow 100	90 \Rightarrow 95 \Rightarrow 100	0.0026
Method of randomization	15 \Rightarrow 25 \Rightarrow 0	0 \Rightarrow 35* \Rightarrow 60	5 \Rightarrow 15 \Rightarrow 80	10 \Rightarrow 20 \Rightarrow 60	0.008
Blinded assessor	35 \Rightarrow 55 \Rightarrow 40	35 \Rightarrow 80 \Rightarrow 60	35 \Rightarrow 65 \Rightarrow 100	65 \Rightarrow 70 \Rightarrow 80	0.0024
Adverse events	25 \Rightarrow 70 \Rightarrow 60	35 \Rightarrow 70 \Rightarrow 80	30 \Rightarrow 65 \Rightarrow 80	30 \Rightarrow 74 \Rightarrow 100	0.0001
Statistical analysis	95 \Rightarrow 100 \Rightarrow 100	95 \Rightarrow 100 \Rightarrow 100	100 \Rightarrow 100 \Rightarrow 100	90 \Rightarrow 100 \Rightarrow 100	0.12
Type I statistical error	58 \Rightarrow 35 \Rightarrow 80	50 \Rightarrow 74 \Rightarrow 40	65 \Rightarrow 50 \Rightarrow 60	55 \Rightarrow 60 \Rightarrow 80	0.87
Type II statistical error	21 \Rightarrow 15 \Rightarrow 0	20 \Rightarrow 0 \Rightarrow 20	10 \Rightarrow 10 \Rightarrow 0	20 \Rightarrow 15 \Rightarrow 0	0.16

Data for each journal are the percent of trials for which the criterion was present during the periods 1981–1985 \Rightarrow 1991–1995 \Rightarrow 2000.

* Based on pooled scores from the four journals for each criterion between 1981–1985 and 1991–1995 only. † $P < 0.0083$, 1991–1995 value compared with 1981–1985 value for the same journal.

riod. For 7 of the 10 criteria, consent, eligibility, randomization, methods of randomization, blinded assessor, and adverse events, the pooled values increased significantly between 1981–1985 and 1991–1995 (table 1). The reporting of sample size calculations was poor in the periods 1981–1985 and 1991–1995, appearing in fewer than 20% of trials in 1991–1995, but the reporting increased to 55% by 2000 (fig. 2 and table 1). Although the reporting of the method of randomization increased between 1981–1985 and 1991–1995 (fig. 3), the difference reached statistical significance for only one journal ($P < 0.0083$). By 2000, 75% of the trials in three of the four journals reported the method of randomization. In the fourth journal, none of the trials reported this criterion. Statistical analysis was reported in most trials and did not change significantly between 1981–1985 and 1991–1995. The risk of a type I statistical error occurred in approximately 50% of the trials during all three periods (fig. 4). The risk of a type II statistical error was

present in approximately 10% of trials (table 1) and did not change significantly between 1981–1985 and 2000.

Discussion

We evaluated the quality of study design in clinical trials published in four leading anesthesia journals between 1981 and 2000. The overall quality of the trials was strikingly similar among the four journals within each period, a finding that is consistent with preliminary data previously reported (Alex Mathieu, M.D., Department of Clinical Epidemiology and Biostatistics, University of Cincinnati, Cincinnati, OH, written communication). The mean analysis scores for the four journals increased approximately 20% between successive periods, an increase that exceeded that reported in other

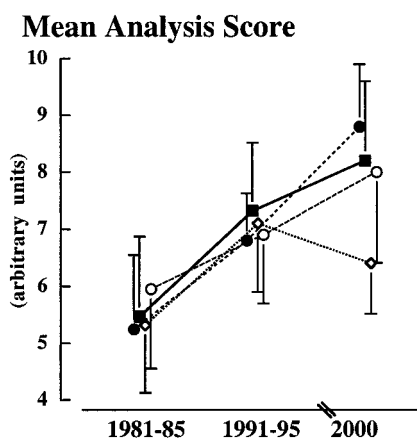


Fig. 1. Mean analysis scores for four leading journals in anesthesia during 1981–1985, 1991–1995, and January to July 2000. The scores for the four journals within each period were similar, but the scores increased from 5.5 ± 1.4 to 7.0 ± 1.1 between 1981–1985 and 1991–1995 ($P < 0.00001$) and to 7.8 ± 1.5 in 2000. The symbols represent the mean scores for the four journals evaluated in each period.

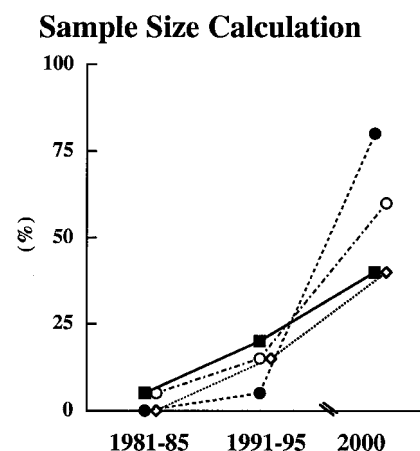


Fig. 2. Sample size calculation was poorly reported in all journals during all three periods. When the rate of reporting the sample size calculation in the four journals was pooled for each period, reporting of the sample size calculation increased between 1981–1985 and 1991–1995 ($P < 0.0174$). The reporting of sample size calculations in individual journals did not change among the three periods. The symbols represent the mean percentages of trials that reported sample size calculation for the four journals evaluated in each period.

Method of Randomization

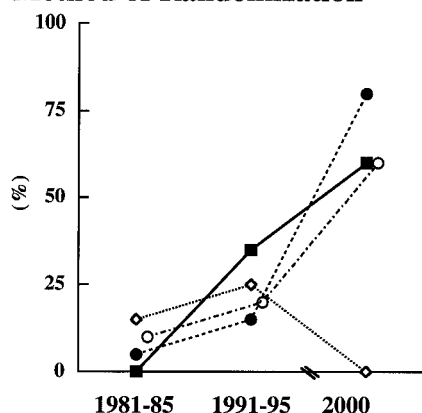


Fig. 3. Method of randomization was poorly reported in all four journals during the three periods, although the reporting increased from 0% in 1981–1985 to 35% in 1991–1995 in one journal ($P < 0.0083$). When the rate of reporting the method of randomization for the journals was pooled within each period, the rate increased significantly from 1981–1985 to 1991–1995 ($P < 0.008$). The symbols represent the mean percentages of trials that reported the method of randomization for the four journals evaluated in each period.

medical subspecialties.^{10–14} The mean analysis score of clinical trials from the four anesthesia journals in 2000, 7.8 out of a maximum value of 10, is encouraging, but with only 30% of the trials scoring 9 or greater of a maximum score of 10, deficiencies remain in the majority of published clinical trials. Although the percentage of trials scoring 9 or greater out of 10 in 2000 is threefold greater than the 10% reported in 1991–1995, the overall quality seems lacking. Greater emphasis on clinical study design is required if we are to remain a competitive medical research subspecialty.

Sample size calculation is frequently omitted from clinical trials.^{14,16–19} Reviews of trials published in journals from medicine, surgery, and family practice have reported low rates, between 5% and 52%, for sample size calculation^{3,10–14,17–21} that are consistent with the 12% reporting rate for sample size calculation during the 1991–1995 period in this study. Although other specialties have not reported comparable data for sample size calculations in the past 5 yr, our finding of a fourfold increase in the reporting of this criterion suggests that sample size calculation is more widely included in the study design of clinical trials in the anesthesia literature than it has been in the past two decades. Nonetheless, that 45% of trials in 2000 did not report a sample size calculation reflects poorly on the scientific, ethical, and fiscal aspects of study design in clinical anesthesia. Scientifically, a sample size calculation is essential because it provides a reasonable estimate of the number of patients needed to reject the null hypothesis if it were truly false, thereby minimizing the risk of type II statistical error.¹⁶ Ethical concerns mandate that a trial should enroll the minimum number of patients required to reject the null hypothesis. The basis for the ethical con-

cern is that a minimum number of patients should be exposed to the potential harms associated with randomizing their care. To address this concern, most ethics committees now require investigators to justify their sample size before ethics committee approval is granted. Finally, the sample size calculation dictates fiscal implications, including the number of patients that should be enrolled, the duration of the study, and the overall study cost. Failure to include a sample size calculation may result in excessive fiscal costs due to overenrollment of patients.

Randomization, a strategy that is used to minimize bias in clinical trials,^{6,17,22,23} consists of two distinct operations: random allocation and an appropriate method of randomization. Random allocation refers to the assignment of patients to treatments such that the chance of receiving any one treatment is the same for all comparative treatments. Randomization sequences that provide an appropriate method of randomization are based on random number tables, computer programs, or any other technique for which the chance that any single treatment is assigned to a patient is the same for all comparable treatments. Arbitrarily assigning patients to treatments, termed nonrandomized allocation, may introduce bias by interfering with or manipulating the treatment assignment.^{22–24} Flipping a coin or use of birth dates or hospital record numbers are acceptable methods of randomization; however, they are unacceptable methods for allocating patients to treatments because the investigator could be unblinded to the treatment assignment.^{6,17,22–24} These techniques are termed non-concealed randomization. Trials in which the allocation assignment is not concealed are more likely to lead to exaggerated treatment effects, resulting in more trials with positive outcomes than trials in which the treat-

Type 1 Statistical Error

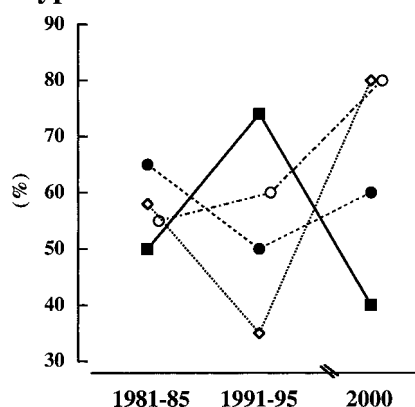


Fig. 4. Type I statistical error was present in almost 50% of all trials evaluated during the three periods. The risk of type I statistical error neither changed significantly for each journal between 1981–1985 and 1991–1995 nor changed significantly when the data from the four journals were pooled during the same periods. The symbols represent the mean percentages of trials that reported type I statistical error for the four journals evaluated in each period.

ment is concealed.^{4-6,20,21,25} Some authors have suggested that concealed allocation, an issue that was not evaluated in the current study, may be more important in minimizing bias than the actual randomization.²¹ The results of this study indicate that almost all of the trials reviewed reported random allocation of the treatments.

Reporting rates for the method of randomization of 7.5% in 1981-1985 and 26% in 1991-1995 in this study and the results of another study²⁶ suggest that the anesthetic literature has not kept pace with the 16-50% incidence reported in medical and surgical specialties during the same intervals.^{3,7,10,17,18,21,23} In 2000, the method of randomization increased to 50% of the clinical trials reviewed, with marked variability among the journals. Comparable data for this criterion from other medical specialties in 2000 are not available. Although omission of the method of randomization does not necessarily indicate that an inappropriate method of randomization was used, it does raise concern regarding possible bias in the data. This criterion must be addressed in future trials.

Blind assessment of outcome is another strategy that minimizes bias in clinical trials. In this review, we scored a trial as blinded if the methodology was described as double-blinded. Implicit in this description is the notion that the patient as well as the observer were blinded to the treatment assignment. We did not require that the anesthesiologist be blinded to the treatment, provided that he or she did not determine the treatment, that he or she could not influence the outcome variables, and that he or she did not assess the outcome variables. Ideally, all participants who could potentially influence the results, including those who observe, record, or interpret any outcomes, should be blinded to the treatment assignment. In 2000, 70% of the trials indicated that the outcome variables were assessed by an individual who was blinded to the treatment assignment. Although this is almost double the rate in 1981-1985, it falls short of the 100% expectation for quality clinical trials. For circumstances in which blinding of the treatment is not feasible, at the very least, the investigators should ensure that the observers are blinded to the study hypotheses.

A key area of study design that showed little improvement and was addressed infrequently was the risk of a type I statistical error. Type I statistical errors occur when multiple comparisons are performed between treatments without statistical compensation.²⁷ This may lead to exaggerated treatment effects (*i.e.*, falsely rejecting the null hypothesis). Using our threshold of 10 between-group comparisons to define the level beyond which the risk of a type I statistical error may become substantive, approximately 65% of the trials in 2000 were at risk for this error. Type I statistical errors are best controlled by prevention: identify the primary outcome variable, limit the number of variables to be eval-

uated, and minimize the between-group comparisons. If multiple between-group comparisons must be performed, then techniques such as multivariate analysis of variance or the Bonferroni *t* test must be used.²⁷

Our definition of a type I statistical error may be critiqued on two accounts. First, we arbitrarily defined the threshold for a type I statistical error as more than 10 between-group comparisons. No published consensus exists for the maximum number of between-group comparisons after which compensation for a type I error is appropriate. However, many would consider 10 between-group comparisons to be excessive and would recommend compensating for a type I error after even fewer comparisons. In this case, the risk of a type I error in our study would have increased. Second, we included all between-group comparisons, whether primary or secondary outcome variables or demographic variables. Some investigators consider only primary outcome variables and the comparisons that relate directly to that variable to contribute to type I errors. Had we accepted this latter threshold for type I errors, then the frequency of reporting this error would have been reduced by as much as 40%. Irrespective of the accepted threshold, to curb exaggerated treatment outcomes, we recommend that the editorial boards adopt criteria for diagnosing and implementing compensation for type I statistical errors in the outcome variable.

One additional possible limitation of this study was our modification of the criteria for evaluating the quality of trials by DerSimonian *et al.*² We modified the criteria to improve the applicability and relevance of the criteria to clinical trials in anesthesiology.² Specifically, criteria from DerSimonian *et al.*,² such as "admission before allocation," "patients' blindness," and "lost to follow-up," have limited relevance to clinical trials in anesthesia. Furthermore, we modified the "statistical methods" by addressing type I and II statistical errors individually. Although the analysis scores based on the 10 criteria have not been validated, it is our contention that they provide a reasonable measure of the overall quality of clinical trials in anesthesia.

In summary, we found that trials published in four major anesthesia journals between 1981 and 2000 were of similar quality. The overall quality of the study design (mean analysis scores) increased approximately 25% between 1981-1985 and 1991-1995 but only 10% between 1991-1995 and the first 6 months of 2000. Nonetheless, serious deficiencies in study design remain, including sample size calculation, method of randomization, blinded assessors, and consideration of type I statistical errors. To remain a viable and competitive clinical-research subspecialty, it is incumbent on clinical investigators in our specialty as well as the editorial boards to acknowledge the deficiencies and support minimal standards for study design to ensure high-quality clinical trials in anesthesia.

The authors thank Ms. Nancy Sikich (Department of Anaesthesia, Hospital for Sick Children, Toronto, Ontario, Canada) for her assistance in completing this study.

References

- Guyatt GH, Cook DJ, Sackett DL, Eckman M, Pauker S: Grades of recommendation for antithrombotic agents. *Chest* 1998; 114:441S-4S
- DerSimonian R, Charette LJ, McPeck B, Mosteller F: Reporting on methods in clinical trials. *N Engl J Med* 1982; 306:1332-7
- Emerson JD, McPeck B, Mosteller F: Reporting clinical trials in general surgical journals. *Surgery* 1984; 95:572-9
- Schulz KF: Subverting randomization in controlled trials. *JAMA* 1995; 274:1456-8
- Schulz KF, Chalmers I, Hayes RJ, Altman DG: Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-12
- Chalmers TC, Celano P, Sacks HS, Smith H: Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309:1358-61
- Mosteller F, Gilbert JP, McPeck B: Reporting standards and research strategies for controlled trials: agenda for the editor. *Controlled Clin Trials* 1980; 1:37-58
- The Standards of Reporting Trials Group: A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272:1926-31
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF: Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA* 1996; 276:637-9
- Liberati A, Himmel HN, Chalmers TC: A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; 4:942-51
- Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC: An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Controlled Clin Trials* 1990; 11:339-52
- Sonis J, Joines J: The quality of clinical trials published in the *Journal of Family Practice*, 1974-91. *J Fam Pract* 1994; 39:225-35
- Solomon MJ, McLeod RS: Surgery and the randomized controlled trial: Past, present and future. *Med J Australia* 1998; 169:380-3
- Ah-See KW, Molony NC: A qualitative assessment of randomized controlled trials in otolaryngology. *J Laryngol Otol* 1998; 112:460-3
- Clarke M, Chalmers I: Discussion sections in reports of controlled trials published in general medical journals: Islands in search of continents? *JAMA* 1998; 280:280-2
- 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 control trial: survey of 71 "negative" trials. *N Engl J Med* 1978; 299:690-4
- Pocock SJ, Hughes MD, Lee RJ: Statistical problems in reporting of clinical trials: A survey of three medical journals. *N Engl J Med* 1987; 317:426-32
- Altman DG, Dore CJ: Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335:149-53
- Slim K, Bousquet J, Kwiatkowski F, Pezet D, Chipponi J: Analysis of randomized controlled trials in laparoscopic surgery. *Br J Surg* 1997; 84:610-4
- Gardner MJ, Bond J: An exploratory study of statistical assessment of papers published in the *British Medical Journal*. *JAMA* 1990; 263:1355-7
- Schulz KF, Chalmers I, Grimes DA, Altman DG: Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994; 272:125-8
- Altman DG: Randomisation: Essential for reducing bias (letter). *Br Med J* 1991; 302:1481-2
- Stewart LA, Parmar MKB: Bias in the analysis and reporting of randomized controlled trials. *Intl J Tech Assess Health Care* 1996; 12:264-75
- Browner WS: Clinical research: A simple recipe for doing it well. *ANESTHESIOLOGY* 1994; 80:923-8
- Peto R: Why do we need systematic overviews of randomized trials? *Stat Med* 1987; 6:233-44
- Avram MJ, Shanks CA, Dykes MHM, Ronai AK, Stiers WM: Statistical methods in anesthesia articles: An evaluation of two American journals during two six-month periods. *Anesth Analg* 1985; 64:607-11
- Glantz SA: *Primer of Biostatistics*, 4th edition. New York, McGraw-Hill, 1997, pp 86-8