

---

# Quality and retrieval of obstetrical anaesthesia randomized controlled trials

Jennifer S Bender BSC,<sup>1</sup> Stephen H Halpern MD, MSc, FRCPC,<sup>1,2</sup> Molly Thangaroopan BSC,<sup>1</sup> Alejandro R Jadad MD, DPHIL,<sup>3</sup> Arne Ohlsson MD, MSc, FRCPC.<sup>2,4</sup>

---

**Purpose:** Randomized controlled trials (RCTs) are suitable for meta-analysis and systematic reviews provided they are of high quality and are easy to retrieve. We determined these attributes of RCTs in obstetrical anaesthesia in a sample of available journals that are indexed in MEDLINE.

**Source:** Randomized controlled trials published between January 1985 and December 1994 in seven anaesthetic and three obstetric journals were identified by a MEDLINE search, and by handsearch of the same journals. Each RCT was assigned a quality score by a blinded rater using a reliable and validated scale. The quality of each RCT was described and compared over time and by journal. The comparative yield of MEDLINE and handsearch was evaluated.

**Principal findings:** Three hundred and forty RCTs were retrieved by MEDLINE and handsearch. Two hundred and twenty seven (65%) were identified by the MEDLINE search and 333 by the handsearch (98%). The median quality score was 3/5. There was no difference in score over time. *Anesthesiology* had the highest median score. *Anaesthesia* had the lowest ( $P < 0.05$ ).

**Conclusions:** Care must be taken when reviewing obstetrical anaesthesia research. Strategies in addition to a MEDLINE search must be used to identify RCTs since more than one third were missed using MEDLINE alone. Poor quality RCTs are more likely to be biased in favour of a new treatment. Therefore, to increase the validity of reviews sensitivity analyses based on quality should be done.

**Objectif :** Les épreuves aléatoires avec contrôles (EAC) se prêtent bien à la méta-analyse et aux survols systématiques en autant qu'ils sont de bonne qualité et faciles à récupérer. Les auteurs ont déterminé les caractéristiques des EAC en anesthésie obstétricale sur un échantillon de revues indexées dans MEDLINE.

**Sources :** Les épreuves aléatoires avec contrôles publiées entre janvier 1985 et décembre 1994 dans sept revues d'anesthésie et trois revues d'obstétrique étaient identifiées au cours d'une recherche dans MEDLINE et par une recherche visuelle des mêmes revues. Un évaluateur neutre assignait un score de qualité à l'aide d'une échelle fiable et validée. La qualité de chaque EAC était vérifiée et comparée relativement au moment de sa parution et de son origine. La compilation obtenue grâce à MEDLINE était comparée au recueil visuel.

**Constations principales :** Il faut être prudent quand on analyse la recherche en anesthésie obstétricale. Comme plus du tiers des EAC a été omis dans MEDLINE, on doit aussi rechercher ailleurs. Les EAC de qualité médiocre sont susceptibles d'afficher un biais en faveur d'un nouveau traitement. Dès lors, pour augmenter la validité des analyses des articles de révision, il faut effectuer des analyses de sensibilité basées sur la qualité.

From the Department of Anaesthesia<sup>1</sup>, Maternal, Infant and Reproductive Health Research Unit<sup>2</sup>, Department of Newborn and Developmental Paediatrics<sup>4</sup>, Women's College Hospital, 76 Grenville Street Toronto, Ontario, Canada M5S 1B2, and the Department of Clinical Epidemiology and Biostatistics, McMaster University<sup>3</sup>  
J.S. Bender was supported by a summer studentship from the Medical Research Council of Canada. The work was presented in part at the International Anesthesia Research Society meeting in Washington DC March 9, 1996.

Address correspondence to: Dr. Stephen Halpern<sup>1</sup>. TEL 416-323-6269; FAX 416-323-6307; E-mail halpern@fm.net

Accepted for publication September 22, 1996.

**R**IGOROUS reviews of research are becoming increasingly important to guide medical decisions and health policy formulation. Reviews that are not based on all the available evidence, or on poor quality evidence, might lead to erroneous conclusions.<sup>1</sup> Meta-analysis, or quantitative review, is a technique that statistically combines the outcomes of randomised controlled trials with the goal of providing a more precise estimate of the treatment effect.<sup>2</sup> This approach may be helpful in resolving issues in obstetrical anaesthesia if the RCTs produce inconclusive or contradictory results.

A computerized search of the database MEDLINE is a commonly used method to identify articles of interest. However these searches are limited by imperfect indexing and search algorithms, and by the incomplete information in the database.<sup>2,3</sup> A second source of bias that may invalidate a meta-analysis is the inclusion of RCTs of low quality that may overestimate the treatment effect in the experimental group.<sup>4</sup>

A recently developed quality scale uses three items directly related to the control of bias; randomization, double-blinding and a full description of withdrawals, to assign a score to an RCT. The scale has demonstrated reliability and validity and can be used by people with various training to estimate rapidly and easily the quality of an RCT.<sup>5</sup>

The purpose of this study was to determine how many RCTs have been published between 1985 and 1994 in commonly available non-subspecialty journals and to determine their quality using a recently available tool. An additional objective was to determine the proportion of RCTs which would be retrieved using a broad MEDLINE search strategy compared to a hand search.

**Methods**

*Articles*

RCTs in obstetrical anaesthesia were collected. These were identified by a MEDLINE search using two search strategies. The first strategy used the following MeSH headings: ANESTHESIA; PREGNANCY; and limited to HUMAN, ENGLISH, and RANDOMIZED CONTROLLED TRIALS. Since the term "RANDOMIZED CONTROLLED TRIALS" first appeared as a separate search term in 1991, the term RANDOM ALLOCATION was used for earlier trials. The second strategy included the first but also included a text term search (title and abstract) with the following terms: OBSTETRICS, PREGNANCY, RANDOMIZED, ANESTHESIA, and ANAESTHESIA. The following truncated text terms were also used "ANESTH:",

"RANDOM:", "OBSTET:" and "ANAESTH:" to ensure the broadest possible search. We included trials published between January 1985 and December 1994 in the following non-subspecialty anaesthetic journals; *Acta Anaesthesiologica Scandinavica*, *Anaesthesia*, *Anaesthesia and Intensive Care*, *Anesthesia and Analgesia*, *Anesthesiology*, *British Journal of Anaesthesia*, *Canadian Journal of Anaesthesia*. In addition, RCTs in three obstetrical journals; *British Journal of Obstetrics and Gynaecology*, *American Journal of Obstetrics and Gynecology*, and *Obstetrics and Gynecology* were included. We excluded abstracts, letters to the editor, and RCTs concerning women in the first and second trimester of pregnancy. Journals were chosen on the basis of availability in the local library collection and MEDLINE indexing. The same journals were then hand searched for additional RCTs. The search was done by researchers who were blinded to the outcome of the MEDLINE searches. The number of RCTs found by the MEDLINE searches were then compared to the number found by hand.

*Quality Scale*

A recently available reliable and validated scale was chosen to rate the retrieved articles.<sup>5</sup> This five point scale has the following components: randomization (maximum two points), double-blinding (maximum two points), and description of withdrawals or drop-outs (maximum one point). These items were specifically chosen because of empiric evidence that failure of any of these could independently lead to a bias towards the treatment effect. The instrument is described in Figure 1.

**Instrument for Measuring Quality of RCTs**

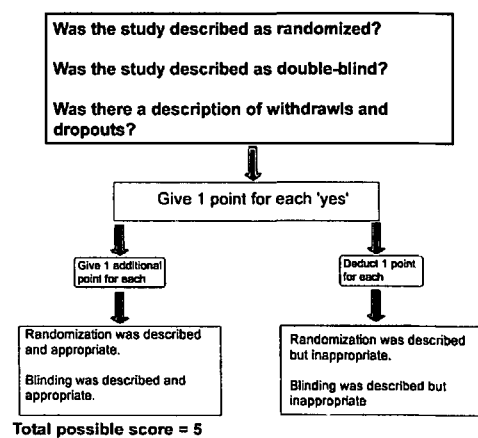


FIGURE 1 Instrument for measuring quality scores.

*Inter-rater agreement*

Before the RCTs were assigned a quality score, the scale was tested to determine the inter-rater reliability. The raters were first trained on the use of the scale. Then, 10 RCTs (not in the obstetrical anaesthesia data set) were chosen randomly from the anaesthesia literature. The papers were blinded as to journal, author, date and acknowledgments. The 10 RCTs were then distributed to each of three trained raters, who assigned scores using the scale described above. Two of these raters were medical students (JB, MT) and one was an anaesthetist (SH). An intraclass correlation coefficient<sup>6</sup> was calculated from the results to determine the agreement of the scale between raters. Differences among raters were discussed and a consensus was reached. The agreement was then tested on a second set of RCTs. An intraclass correlation coefficient of >0.80 was considered high enough to allow each subsequent RCT to be rated by one researcher. The trials from different journals were given to each researcher randomly.

*Quality Assessment of the Obstetrical Anaesthesia Randomized Controlled Trials*

The RCTs that were retrieved using the MEDLINE and hand search were distributed to the three raters with the journal name, authors, affiliations, date of publications and acknowledgements removed. Each trial was scored for quality by one of the raters. Each rater scored between 100 and 120 studies. Comparisons regarding quality were undertaken by journal using the Kuskal-Wallis test. Regression analysis was used to search for trends in quality over time. The RCTs retrieved by hand were compared with those identified by MEDLINE using the Mann-Whitney test. A P value of < 0.05 was considered statistically significant.

**Results**

*Retrieval of RCTs*

A total of 333 RCTs were found by hand searching the journals. Using the first search strategy with MEDLINE MeSH terms only, 178 RCTs were identified by the MEDLINE search, 157 additional RCTs were found by the hand search. Two of those found by MEDLINE were discarded because they referred to first trimester interventions. The second approach, using text terms, identified 277 studies. Of these, 62 were discarded for the following reasons: three were letters to the editor, 27 concerned minor gynaecological procedures (including terminations of pregnancy), 13 compared obstetric interventions, four were animal studies, and 12 studied the wrong population

(eg healthy volunteers, elderly patients), two were non-randomized, and one was a case series. Interestingly, seven RCTs were found using the second search that were missed by the hand search.

The total number of RCTs available, therefore, were 340 (333+7). Of these, 221 were identified using the combined search strategies of MeSH headings and text terms—65% of the total. Because the obstetrical journals contained only 16 RCTs, they were considered together for the purpose of analysis.

*Inter-rater agreement for the quality scale*

The intraclass correlation coefficient calculated between the three raters was 0.85.

*Quality of the research*

The distribution of scores is shown in Figure 2. The number of RCTs and scores of the components of the quality scores are shown in the Table.

The median score was 3 with a range between 0 and 5. When the scores were analyzed by rank (using the Kuskal-Wallis test) *Anesthesiology* had the highest score and *Anaesthesia* had the lowest. For graphical purposes, the ranks obtained from the Kuskal-Wallis test were normalized (transformed to a distribution

**QUALITY SCORES**

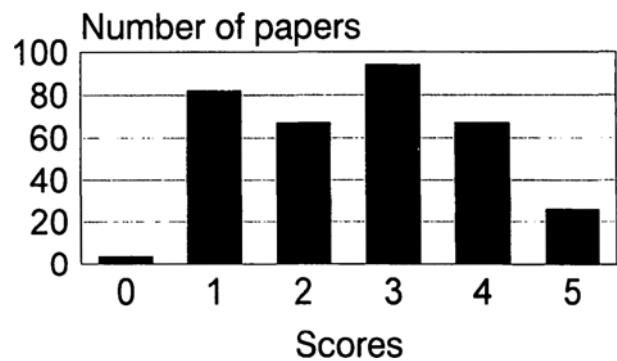


FIGURE 2 Histogram showing the number of papers achieving each score.

TABLE Components of the Quality Score  
Scores are shown as n (%)

DESCRIPTOR	SCORES		
	0	1	2
Randomization	14 (4)	247 (73)	79 (23)
Blinding	142 (42)	40 (12)	158 (46)
Withdrawals	209 (61)	131 (39)	-

with a mean of 0 and a standard deviation of 1.0). The rank scores of both *Anesthesiology* and *Anaesthesia* were statistically different from the median (Figure 3). There was no significant difference between the quality of the RCTs identified by MEDLINE and those retrieved by the hand search. Although there was a difference between years, the slope of the regression line over time was 0.39 and not different from 0. A graph of the normalized rank of papers over the ten year period is found in Figure 4.

### Quality Scores by Journal

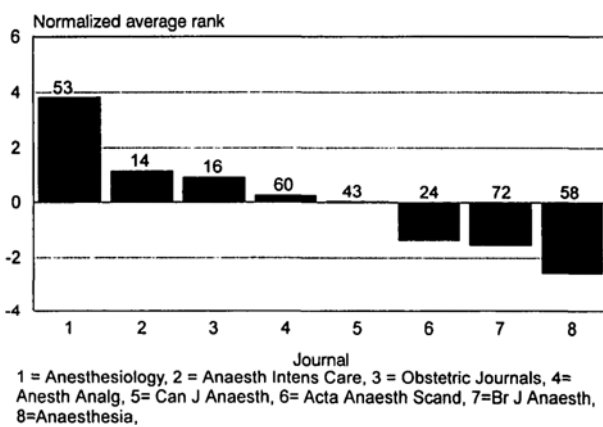


FIGURE 3 Normalized average quality scores are shown by journal. Note that journals ranking  $>1.96$  or  $<-1.96$  are statistically different from the mean. The number of papers in each journal is shown above each bar.

### Quality Scores by Year

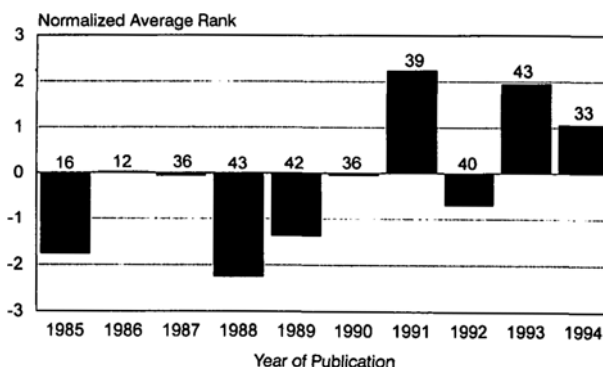


FIGURE 4 Normalized average quality scores shown by year. Note that years ranking greater than 1.96 and less than  $-1.96$  are statistically different from the mean. The number of papers in each year is shown above each bar.

### Discussion

The broad MEDLINE retrieval strategy used in this study failed to identify 35% of the relevant RCTs present in the selected journals. This result is consistent with others who have compared the retrieval rate from MEDLINE with another "gold standard" for total number of trials.<sup>7</sup> There may be several reasons for the failure to identify relevant papers using MEDLINE. MEDLINE consists of approximately 4100 journals reporting research in fields related to medicine.<sup>2</sup> This is only about  $\frac{1}{3}$  of the medical journals currently in print. Although the inclusion of journals in MEDLINE is reassessed periodically, many new journals appear monthly and these are not incorporated into the MEDLINE database. Since we confined our search to MEDLINE indexed journals, this was not a factor in the failure to find all the papers in this case. Other problems include faulty indexing, and faulty search algorithms. Although our algorithm was simple, it was meant to be all inclusive. Others have also noted that known articles are often missed even when an "expert" searcher designs the algorithm. Dickersin *et al.* noted that of 88 studies registered in the Perinatal Database of Clinical trials on the topic "Neonatal hyperbilirubinemia," 81 were indexed in MEDLINE. Of these only 28 could be retrieved by an "expert" searcher and only 11 were retrieved by an "amateur."<sup>7</sup> When 15 other "best sensitivity" MEDLINE searches were reviewed in the journals indexed by MEDLINE, the mean sensitivity was 77% with a range of 32 to 91%, depending on the topic.<sup>7</sup>

It is necessary to use a more extensive approach to retrieving articles than a computerized search of MEDLINE if the application requires all of the relevant articles. Other databases (EMBASE, The Cochrane database of systematic reviews,<sup>8</sup> subspecialty databases, etc), personal files and references found in retrieved articles are excellent sources of additional material. It is reassuring that the quality of those RCTs retrieved was not different from those that were missed. Therefore, if a meta-analysis were to be performed there would be a loss of power to detect relevant clinical outcomes but one would not expect bias to be introduced because of poor quality. It is not known, however, whether there is a systematic difference in effect size or direction between studies retrieved by MEDLINE and those that are not because of other variables (retrieval bias).

The quality scores assigned in this study indicate that although some trials do properly address and report the three items in the scale, many did not. Improper randomization consists of schemes that do not generate an unpredictable assignment sequence or

allow the investigator to know the treatment group before the patient is recruited into the study. Examples of these include randomization by hospital number, day of the week, or "preference of the anaesthetist." These may lead to selection bias that favours the "experimental" treatment by as much as 41%.<sup>4</sup> Most RCTs stated that the groups were chosen "randomly" but did not detail the randomization scheme (73%—see Table). Schulz *et al.* found that in their sample of obstetrical RCTs, the odds ratios were exaggerated in favour of the experimental group by 30% when the method of randomization was not reported.<sup>4</sup>

Double-blinding was either not reported or inadequately reported in 46% of our sample (Table). Inadequate blinding may lead to a bias in favour of the experimental group of 17%.<sup>4</sup>

Withdrawals from the study after randomization were only documented in 39% of papers. It is important for withdrawals to be described for several reasons. The withdrawals may have occurred because of the treatment and therefore some treatment failures were excluded from the analysis. This would clearly bias the results in favour of the alternative group. Additionally, if there were many withdrawals compared with the number of subjects in the study, there may be a threat to randomization.

There was no change in the quality of RCTs over time. This may be because some years contained too few RCTs to analyze. In addition, some trials may not be reported for some years after completion. It would be interesting to follow the trend in quality forward to determine if improvement in quality over time could be demonstrated.

The quality of RCTs differs among journals. This may be due to differences in the rigor of the peer review process and instructions given to authors<sup>9</sup> and reviewers. Recently, some journals have ensured that each trial undergoes a statistical assessment before being accepted for publication.<sup>10</sup> Another approach may be to give explicit instructions to authors to report their methodology fully so that readers can assess whether or not the best possible study quality was employed.<sup>11</sup> Similarly, reviewers should consider markers of quality and the possibility of bias due to study design when recommending acceptance or rejection of RCTs.

Sensitivity analysis is a method of evaluating the stability of conclusions based on assumptions made in the analysis.<sup>2</sup> Sensitivity analysis could be applied to RCTs used for meta-analysis based on the quality of the study to determine if quality has an effect on the conclusions. If including poor quality studies has an effect on the outcome of the analysis, an argument could then be made to exclude these trials from the analysis or, perhaps more appropriately, to advise cau-

tious interpretation of the results. Therefore, to increase the reliability of a meta-analysis, sensitivity analyses based on quality should be performed.

In summary, a broad MEDLINE search strategy seeking RCTs in obstetrical anaesthesia failed to recover 35% of the relevant studies. Many of those found were inadequately reported and likely to be flawed and biased in favour of the experimental group because of inadequate study design. Care must be exercised when including studies of low quality in meta-analyses and sensitivity analysis should be performed to determine the effect of quality on outcome. The peer review process should ensure rigorous assessment and adequate reporting of basic study design quality before publication of an article is permitted.

## References

- 1 Moher D, Olkin I. Meta-analysis of randomized controlled trials. A concern for standards. *JAMA* 1995; 274: 1962-3.
- 2 Petitti DB. Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis. *Methods for Quantitative Synthesis in Medicine*. New York: Oxford University Press, 1994.
- 3 Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286-91.
- 4 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.
- 5 Jadad AR, Moore RA, Carroll D, *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*. 1996; 17: 1-12.
- 6 Streiner DL, Norman GR. *Health Measurement Scales. A Practical Guide to their Development and Use*, 2nd ed. Toronto: Oxford University Press, 1995.
- 7 Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Controlled Clinical Trials* 1985; 6: 306-17.
- 8 Bero L, Rennie D. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; 274: 1935-8.
- 9 *The Standards of Reporting Trials Group*. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272: 1926-31.
- 10 Pitkin RM, Burmeister LF. Routine statistical screening revisited (Editorial). *Obstet Gynecol* 1995; 86: 124-5.
- 11 Begg C, Cho M, Eastwood S, *et al.* Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA* 1996; 276: 637-9.

## ADDENDUM

References to each article retrieved are available from the authors upon request.