Special Article

The literature of anaesthesia: what are we learning?

Peter G. Duncan MD FRCPC,
Marsha M. Cohen MSC MD MHSC FRCPC

In an effort to identify the types of articles published in anaesthesia literature, a stratified random sample of articles published in North America between 1977 and 1986 was analyzed (N = 571). Human studies constituted 63 per cent of the total, with case reports and case series constituting over half. Study designs classed as descriptive in nature were remarkably rare in the anaesthesia literature, with prevalence and case-control studies constituting 0.8 and 3.3 per cent of the total respectively. Cohort studies (7.8 per cent), non-randomized intervention studies (12.8 per cent), and randomized controlled trials (17.8 per cent) were more numerous, but many suffered major contamination of experimental design. Frequently identified concerns in assessing the applicability of a given study to general anaesthetic practice were a bias induced by selection of the study subjects, application of the results from tertiary care hospitals to community hospitals, and contamination of the study protocol. These factors were identified as present in the majority of articles. The results suggest that growth of the specialty of anaesthesia is constrained by the narrow spectrum of study designs, as well as major problems affecting generalizability of the published results.

Key words

ANAESTHESIA; education, continuing; PUBLICATIONS; anaesthesia, research.

From the Department of Anaesthesia, University of Saskatchewan, Saskatoon, Saskatchewan and the Department of Community Health Sciences, University of Manitoba (MMC), Winnipeg, Manitoba

Address correspondence to: Dr. Peter Duncan, Department of Anaesthesia, University Hospital, Saskatoon, Saskatchewan, S7N 0X0.

The majority of adult learning is accomplished to some extent by use of the written word. As such, when considering our role as contemporary anaesthesiologists, we "are what we read." The question must be asked whether the currently available literature, reflective of the study methods currently used in anaesthesia research, is appropriate to individual practice development. Incorporation into practice of the information contained in a given piece of literature is dependent upon the study design and the generalizability of the results – to extrapolate a study's results to one's individual practice is hazardous unless that study has met clearly defined criteria. In addition, development of the specialty as a science is dependent upon maintaining a broad scope of research done by individuals with diverse views, ideas, and techniques – excessive reliance on confined patterns of investigation may ultimately be a hindrance to anaesthesia achieving its rightful place in the health care sphere. The purpose of this study was to examine a sample of the North American anaesthesia literature with respect to study design and factors which may influence effective development of individual anaesthesia practice.

Methods

The study journals consisted of the three main North American publications offering original research articles in the specialty – Anesthesiology, Anesthesia and Analgesia, and the Canadian Anaesthetists' Society Journal. A stratified random selection process (roll of dice) chose one issue from each year for each of the journals. The years of publication were 1977–1986 inclusive, resulting in 30 issues analyzed. It is noteworthy that during the study period the Canadian Journal and, for two years, Anesthesia and Analgesia, were published bimonthly.

The contents of the selected issues were then grouped according to study type. The classification utilized included:

- 1 Human investigation of any type;
- 2 Animal or laboratory articles;
- 3 Continuing Medical Education or review articles;
- 4 Other articles, including editorial comments, historical presentations, or human interest topics.

TABLE I Study classification

Type	Sample selection pitfalls	Theoretical or actual example	Evaluation criteria
<i>Descriptive</i>			
Case report			
Case series	Pre-selection bias	Chart audit of post-surgical deaths at one hospital	
Prevalence or cross-sectional	Pre-selection bias	Survey of all hospitalized patients to compare the rate of sore throats in surgical and non-surgical patients	Major: Incidence-prevalence bias Diagnostic equivalence bias
Case-control	Representativeness of cases or controls	Cases of non-infectious hepatitis and cases free of hepatitis compared to determine prior exposure to halothane	Nil: Has appropriate controls, no major biases Minor: Has biases, but easily explained or controlled for Major: Major bias not explained, e.g., detection bias, Neyman bias
Cohort-analytical	Unrepresentative sample. Inappropriate comparison group	Pregnant women who did and did not have an anaesthetic (operative procedure) followed to determine rate of congenital anomalies ¹⁸	Nil: Representative sample, little attrition, definitive end points Control for confounding points Minor: Measured attrition
Cohort-descriptive	Unrepresentative sample	All patients receiving an anaesthetic at one hospital for a nine-year period, rate of intraoperative, recovery room and postoperative complications determined in those with and without preoperative medical conditions ¹⁹	Major: Non-random attrition Same as for analytical.
<i>Intervention</i>			
Quasi-experimental (non-random)	Volunteer effect	All patients at one hospital receive halothane and all patients at a second hospital receive ethrane – rate of intraoperative complications is determined	Nil: Has concurrent comparison group, blinding, no intervention Minor: Comparison group either self or historical, lacks blinding Major: Cointervention present, failure to account for confounders
Experimental (Randomized controlled trial).	Too strict inclusion criteria, limited generalizability	Each surgical patient to receive an inhalational agent is randomly assigned to receive either halothane or ethrane. The rate of intraoperative complications is then determined	Nil: Proper randomization, no attrition, blinded Minor: Suspect randomization, no blinding, low attrition Major: Contamination of protocol

The articles pertaining to human investigation were further divided as to methodology (Table I). There are two basic forms of investigation: descriptive and intervention (or experimental). Descriptive studies describe the pattern of disease occurrence in human populations, while in intervention studies the investigator actively intervenes to change one variable, then examines dependent variables (outcomes). For each study type, sources of bias which limit application of the results were also sought.

The human studies were classified as follows:

1 Case reports/series.

This was defined as a single (or multiple) case history usually presented with a review of the literature referable to the course of management. It is the simplest of the descriptive forms of study and merely describes an isolated observation to the readership, using no

comparison data. Chart audits were included in this group.

2 Prevalence or cross-sectional study

In this type of study events were defined as they occur at a discrete point in time in a specified population. Typically a survey is carried out to determine which individuals have the disease of interest, how many have the characteristic which might be affecting the disease outcome, and how many have both or neither. Common problems affecting the generalizability of this type of study include the method of sampling the total population (ie: was the sample truly representative?) and whether the diagnostic criteria for presence of the disease or characteristic were clearly defined. The study must clearly avoid the "incidence-prevalence bias",¹ the former being a description of new events or

cases occurring in a defined population over time, while the latter is the number of events or cases present at one particular point in time.

3 Case-control study

A study of this type takes a group of persons with a given disease state and a group without this disease and looks backward to examine factors which may have had bearing on development of that state.² Selection validity must ensure that the study cases are indeed representative of all persons with the disease, and that the control group indeed represents all control subjects.³ Common contaminants (biases) of this form of study include detection bias (controls not free of disease), anamnestic equivalence (ensuring that both cases and control were subject to equal scrutiny for the presence of antecedent events), and Neyman's bias (disease with a transient history due to death or recovery may be under represented in the case population).^{5,6} Case-control studies were formerly termed retrospective, and are particularly useful in studying rare events.

4 Cohort or incidence study

This form of study (known also as a prospective study) can be divided into two main categories. The descriptive form takes a clearly defined population who will be submitted to a treatment or have a preexisting risk factor but are free of the disease of interest. They are followed over time to observe the rate of development of a certain event, disease or complication. The key elements are the representativeness of the sample and that the population is defined in advance. The analytical form of a cohort study takes a group of subjects with a certain characteristic and follows them forward in time. A suitable comparison group drawn from the same population but without the characteristic under study is similarly followed. Methodological concerns are based on selection of the cohort, the absence of the characteristic in the control group, control of confounding variables (factors which co-exist and may affect the outcome), and the absence of attrition from either group.^{3,6} When attrition does occur, it must be measured and shown not to occur in any definable fashion (for example, subjects may be lost from one group because they all died from the disease of interest, a fact unknown to the investigator). In contrast to experimental studies, the observer does not determine the course of therapy but only measures the extent of exposure to the element in question.

5 Intervention studies

For this type of study, the manoeuvre is under control of the investigator. This form of study may be experimental (randomized controlled trial) or quasi-experimental in nature. In the former, volunteers or patients are randomly assigned to either receive or not

TABLE II Classification of articles

	(n)	% of Total
Human investigation	359	63.0
Laboratory or animal	136	23.7
Continuing medical education	35	6.1
Other (editorial, historical)	41	7.2
	571	100

receive the intervention or treatment of interest, and clearly defined endpoints for improvement stated in advance. In quasi-experimental studies, the study group is subjected to the intervention and a comparison group serves as a control. The comparison group may be an analogous group of individuals, a historical group who did not receive the intervention or treatment, or the same group "before and after." In quasi-experimental designs, the assignment of the intervention is generally by group rather than by individuals. As in any study, problems arise through such factors as cointervention or attrition from the study.⁷ Generalization is dependent upon the selection of study subjects, most often termed a "volunteer bias" (implying those who consent to the study may differ in some critical way from the usual anaesthesia population). A critical element for intervention studies is blinding, either single (the subject) double (the subject plus the evaluator) or triple (the above plus the person undertaking the data analysis).

For a fuller review of how epidemiologic concepts can be applied to anaesthesia see Farrow and Fowkes.⁸

Each of the articles reviewed was classified as to type, the presence of any biases, and whether there might be contamination of the study design based upon the criteria set out in advance in Table I. Initial classification of all articles was done by one author (PGD). When doubt existed consensus agreement was reached by both authors. Interobserver reliability was not measured as part of this study. When the contamination was not felt to influence critically the study results it was rated as minor; major contamination included factors which invalidated the study, thus precluding the incorporation of the results into clinical practice.

No attempt was made to evaluate the statistical analysis deployed by the various studies, the assumption being these have been appropriately reviewed by the respective editorial process.⁹

Results

The articles published in the selected journals showed a predominance of human investigation, as shown in Table II. Animal or laboratory work occupied 23.7 per cent of

TABLE III Analysis of human studies

Study type	Total		Selection bias		Nil	Contamination	
	(n)	(%)	(n)	(%)		Major	Minor
Case reports	132	(36.7)	132	(100)	—	132 (100%)	—
Case series	74	(20.6)	74	(100)	—	74 (100%)	—
Prevalence	3	(0.8)	1	(33.0)	3 (100%)	—	—
Case-control	12	(3.3)	8	(75.0)	2 (16.7%)	6 (50.0%)	4 (25.0%)
Cohort	28	(7.8)	14	(50.0)	19 (67.8%)	1 (3.6%)	8 (44.4%)
Non-random intervention	46	(12.8)	11	(24.0)	6 (13.0%)	8 (17.4%)	32 (69.6%)
Randomized control trial	64	(17.8)	60	(94.0)	26 (40.6%)	19 (29.7%)	19 (29.7%)
Total	359		300	(83.6)	56 (15.6%)	240 (66.9%)	63 (17.5%)

published articles, with small contributions made by continuing medical education, historical, and editorial comments. The percentage distribution, while variable between issues of interest, remained remarkably constant over the years when all three contributing journals were included.

Analysis of the studies done on human subjects is presented in Table III. Most numerous were articles classified as a case series, often with $n = 1$ and published as a case report. Together they comprised 57 per cent of the articles and by their nature lacked a reference comparison group. There were remarkably few studies of prevalence of problems in the sampled anaesthesia literature (three reports only), and in one of these the cases sampled were not representative of the typical anaesthetic population.

Case-control (or retrospective) studies constituted 3.3 per cent of the total articles, and most of these were clearly biased in their ability to represent the majority of the available cases of a given disease. In addition, half of the studies had major problems in their design or lacked controls necessary for true case-control comparison, rendering interpretation difficult.

A commonly used descriptive format might be classified as "incidence" or "cohort" in style, representing 7.8 per cent of the total. However, all but 14 of these studies contained a potential selection bias in the study description — either the cases were highly selected, or potentially differed from the general population in some identified fashion. In this group of studies minor contamination was common, usually in the form of attrition; 19 articles (67.8 per cent) satisfied the criteria necessary to exclude contamination.

Quasi-experimental studies of a non-random type remain popular in anaesthesia. With respect to protocols, almost one article in four utilized healthy volunteers who were not similar to the typical subject receiving anaesthetic care. Minor contamination was common (70 per cent) in these studies in the form of comparison groups of

debatable validity, or lack of blinding. In addition, a further 17.4 per cent of articles had a major contamination of protocol in the form of co-intervention partially affecting results.

Randomized clinical trials were very prevalent in the anaesthesia literature, representing 17.8 per cent of the total. However, their generalizability was often suspect as they represented an artificial situation with respect to subject selection. In addition, only 40 per cent of these studies were free of errors, the remainder demonstrating variable amounts of major or minor contamination.

Discussion

It is obvious that anaesthesia as a specialty has made major achievements in recent years, offering our patients greater safety and comfort as we understand more and more about our craft. However, while we may know much more about the science of anaesthesia, it may be suggested that the literature has become too confined in scope and reductive in reasoning, hampering the growth of the enterprise. If this is in fact true, will we as practitioners simply continue to know more and more about less and less, until we ultimately know everything about nothing at all?

There are many types of articles in the medical literature, not all styles of which are suitable for all specialties or the entire medical readership. However, each particular style has both its advantages and disadvantages. Case reports, for example, serve to suggest "I've had a case like this — have you?" Case series are even more numerous than reflected in the literature, for they are the basis for most morbidity reviews in individual departments. Prevalence studies, when properly performed, describe patients and their problems, and allow for comparisons between different points in place and time. As such, they suggest new ideas to be investigated by other means. True case-control studies have been used infrequently in anaesthesia research. However, this type of study is particularly useful in the study of rare events,

and as such may be very suitable for study of rare anaesthetic complications. When properly done they too are able descriptors of problems and relationships to be followed up in future work. It is indeed unfortunate that these latter two descriptive forms of study have been appearing so seldom in the anaesthesia literature, for they are indeed "hypothesis mills" useful for future reference.

Probably due to the professed need for "scientific purity," as well as the acute nature of the anaesthetic experience, there has been an emphasis on controlled studies of either the incidence (descriptive) or intervention (manipulative) style. At times the differentiation between the two was anything but clear as a given study was assessed. Unquestionably the prospective experiment gives the most controlled protocol and the highest probability of acceptance for publication. However, both types require an idea from which to begin, and thus add little true "new information" to the literature.¹⁰ They are critically necessary to support a causal argument about a given question, but generally add no new answers – only confirming a previous suggestion.

There is a major impediment to the generalization or applicability of the results of these studies revolving around the issue of informed consent and the referral pattern of the hospital doing the study.^{11,12} The practice of anaesthesia does not usually allow sufficient contact with patients to assure consents will always be granted, and hence any group of study subjects will potentially be biased if the "volunteers" are systematically different from those who refuse. In a similar vein the results of treatment rendered in different hospitals may vary, both as regards to who is offered surgery and the success of procedures performed.¹³ The reader of articles arising from tertiary care institutions may well find that techniques and procedures advocated may not be applicable to the community hospital practice with its unique selected population of patients.

Undoubtedly the purest form of clinical research is the randomized clinical trial, and these are amply represented in the reviewed literature. While methodologic contamination was found to be common, the greatest difficulty with such trials again revolves around the issue of generalizability.^{14,15} Most properly, the results of such trials are applicable only to the individuals upon whom the study was performed. To assist in interpretation the method of selection of subjects, and their representativeness of the general population, must be clearly described. While the significance of volunteer bias will be dependent on the variables under consideration, as a general rule extrapolation of results to routine clinical practice must be done with caution.

One final influence affecting generalization of results is the Hawthorne effect – the fact of participation in a study

may give results different from those achieved in general application of a particular therapy.¹⁶ Results obtained with anaesthesia care are highly dependent upon the experience of the anaesthetist – in the hands of an experienced and talented investigator the efficacy of a given technique or technology may be very different from that obtained when used by a large number of practising physicians.

In summary, we have demonstrated that a sampling of North American anaesthesia literature has a rather narrow range of study designs. Those most often used for clinical investigation are "controlled" trials, which are effective in confirming a previously stated hypothesis, or case series, which are neither representative nor scientific. At the same time there is a shortage of prevalence or case-control studies designed to either generate totally new hypotheses or study rare events, perhaps suggesting a relative shortage of innovative ideas and broadening concepts. Indeed, the specialty may be a veritable "prisoner" of its own literature! In addition, while methodologic defects and contaminants are not uncommon with any of the studies reviewed, a major problem rests with generalization of the results to routine practice. Bias in selection of subjects, induced by each of the institution of origin, the investigator, and the patient who volunteers, must always be excluded before new techniques are adopted as proven to be efficacious.¹⁷

References

- 1 Friedman GD. Medical usage and abuse: "prevalence" and "incidence". *Ann Int Med* 1976; 84: 502–4.
- 2 Hayden GF, Kramer MS, Horwitz RI. The case-control study: A practical review for the clinician. *JAMA* 1982; 247: 326–31.
- 3 Feinstein AR, Horwitz RI. Double standards, scientific methods and epidemiologic research. *N Engl J Med* 1982; 307: 1611–17.
- 4 Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. *Am J Med* 1979; 66: 556–64.
- 5 Greenland S, Watson E, Neuta RR. The case-control method in medical care evaluation. *Medical Care* 1981; 19: 872–8.
- 6 Sackett DL. Bias in analytic research. *J Chron Dis* 1979; 32: 51–63.
- 7 Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309: 1358–61.
- 8 Farrow SC, Fowkes FGR. Epidemiological strategies in research in anaesthesia. In *Epidemiology in Anaesthesia*. JN Lunn (Ed). Edward Arnold (Publishers) Ltd., London 1986.

- 9 Feinstein AR. Clinical biostatistics II. Statistics versus science in the design of experiments. *Clin Pharmacol Ther* 1970; 11: 282-92.
- 10 Feinstein AR. An additional basic science for clinical medicine: II. The limitations of randomized trials. *Ann Int Med* 1983; 99: 544-50.
- 11 Altman DG. Statistics and ethics in medical research. *Br Med J* 1980; 281: 1267-9.
- 12 Grad RM. Human experimentation and informed consent. *Can Med Assoc J* 1984; 131: 932-3.
- 13 Chassin MR, Brook RH, Park RE et al. Variations in the use of medical and surgical services by the medicare population. *N Engl J Med* 1986; 314: 285-9.
- 14 Fyfe IM. The randomized clinical trial: panacea or placebo. *Can Med Assoc J* 1984; 131: 1336-9.
- 15 Kramer MS, Shapiro SH. Scientific challenges in the application of randomized trials. *JAMA* 1984; 252: 2739-45.
- 16 Sackett DL. Design, measurement and analysis in clinical trials. In: Platelets, Drugs and Thrombosis. Hirsch J et al. (Eds). Basel: S. Karger. 1975. pp. 219-7.
- 17 McPeck B. Inference, generalizability, and a major change in anesthetic practice. *Anesthesiology* 1987; 66: 723-4.
- 18 Duncan PG, Pope WDB, Cohen MM, Greer N. Risk of anaesthesia and surgery during pregnancy. *Anesthesiology* 1986; 64: 790-4.
- 19 Duncan PG, Cohen MM. Postoperative complications: Factors of significance to anaesthetic practice. *Can J Anaesth* 1987; 34: 2-8.

Résumé

Afin d'identifier le genre d'articles publiés dans la littérature anesthésique, un échantillon randomisé et stratifié des articles publiés en Amérique du Nord entre 1977 et 1986 a été analysé ($N = 571$). Les études cliniques constituaient 63 pour cent du total, avec des histoires de cas et des séries de cas constituant plus que la moitié. Les études descriptives étaient rares alors que les études de prévalence et les études cliniques formaient respectivement 0.8 et 3.3 pour cent du total. Les études de cohortes (7.8 pour cent), des études d'intervention non randomisée (12.8 pour cent) et des études contrôlées et randomisées (17.8 pour cent) étaient plus nombreuses mais plusieurs souffraient de problèmes majeurs de planification expérimentale. Les points fréquemment identifiés dans l'évaluation de l'applicabilité d'une étude donnée à la pratique anesthésique générale étaient: un biais induit par la sélection de l'étude, l'application des résultats d'un centre de soin tertiaire à des hôpitaux communautaires, et des problèmes de protocole de l'étude. Ces problèmes ont été identifiés dans la majorité des articles. Ces résultats suggèrent que la croissance de la spécialité anesthésique est contrainte par un spectre étroit de la planification des études ainsi que par des problèmes majeurs affectant la généralisation des résultats publiés.