

Editorial

Clinical research in anaesthesia; randomized controlled trials or observational studies?

The new millennium has seen a consolidation of the difficulties in pursuing clinical research in anaesthesia. Anaesthesia is rarely seen as a specialty with a high academic profile [1]. Universities have been encouraged to concentrate their research effort on basic sciences, and research in clinical medicine has concentrated on those areas with a large impact on public health. In the UK cancer, cardiovascular disease, mental handicap and HIV/AIDS have been identified as the main areas to support, and as a result, funding for research in anaesthesia and related topics has been more difficult than ever. This is unfortunate, since many questions in anaesthesia remain appropriate subjects for research. Given that clinical research is expensive and time consuming, and that funds are scarce, it behoves us to undertake research in a manner that is as effective as possible. Certainly we should not waste valuable resources on inappropriate research methodologies.

The randomized controlled trial (RCT) remains the most highly regarded methodology in clinical research [2] and indeed has been so for many decades. Such studies, when properly conducted, are most effective for postulating and testing a clinical hypothesis. The use of randomization is particularly important, in that it ensures an unbiased allocation of treatment [3] and allows analysis of the data based on statistical theory on the basis of random sampling [4]. Although techniques such as alternation may, if strictly adhered to, be an effective alternative to randomization, knowledge of the sequential allocation of treatment may lead to bias, and for this reason random number tables were introduced into clinical trials design over 50 yr ago [5]. However, randomization does not ensure the direct comparability of two groups of patients in all aspects, and baseline comparability still has to be both tested and demonstrated. But it does ensure that such differences as may occur do so by chance. Another strength of the RCT is the blinded or masked

nature of the treatment allocation. Either the patient does not know to which treatment group he/she is allocated (single blind), or both the patient and the clinician observing the effects of treatment are blinded to the allocation of treatment [6].

Sometimes it is impossible for the clinician giving the treatment to be unaware of the treatment group allocation, e.g. in a comparison of operative surgery techniques, and in those circumstances both treatment efficacy and all complications should be evaluated by third parties blinded to the treatment received. Sometimes an observer will almost inevitably be aware of the treatment allocation, e.g. in a placebo controlled trial of an intravenous (i.v.) vasodilator in the setting of perioperative blood pressure (BP) control. However, in a controlled trial of two active drug treatments it is frequently possible to ensure proper randomization and a double blind design, e.g. using 'double dummy' techniques thus ensuring a rigorous and effective methodology.

Despite their obvious value, we should not delude ourselves into thinking that every question in clinical medicine will be submitted to the rigours of the RCT. Many have bemoaned the lack of RCTs in assessing the efficacy of alternative medicine techniques [7], despite the enthusiastic advocacy of patient groups and practitioners for these therapies. However, alternative medicine is not the only blind spot in assessing clinical efficacy. The cost and difficulties in organizing research can dissuade workers from RCTs in certain settings particularly where 'off label' prescribing is commonplace and accepted. Paediatric anaesthesia is a particularly good example of this problem [8].

Nonetheless, RCTs are rightly regarded as a powerful investigative tool, and indeed it has been suggested following a large study of non-randomized intervention studies that such studies are almost never free of selection bias, and that non-randomized studies should only be undertaken when RCTs are either infeasible or unethical [9]. Useful observational studies on the effects of treatment (i.e. efficacy) are

Correspondence to: R. O. Feneck. E-mail: rob_feneck@msn.com

Accepted for publication 4 October 2006 EJA 4199

rare, and such studies frequently give different results to randomized trials, but the nature of that difference is often unpredictable [10]. Given the above, why is it that observational studies appear to be increasing almost at the same rate as RCTs are declining?

If we consider the difficulties in setting up a RCT we can identify a number of reasons for this, some more creditable than others. I have already mentioned the cost of a RCT in a cash limited environment. All trials in which patients are allocated a treatment at random rather than under the judgement of a clinician require written informed consent and are subject to the Declaration of Helsinki. The organization and simple paperwork involved in seeking Ethical or Review Board approval for a RCT is increasing in a manner that appears to be exponential.

RCTs themselves are not problem free. Randomization may be difficult in certain settings, and there has been a failure to report negative results, particularly in sponsored trials [11]. Some have also stressed the potential for bias in conducting or reporting sponsored trials [11,12]. In my personal experience over 25 yr, I have been more aware of the failure to report negative findings than evidence of bias, but occasionally evidence emerges of the strains that can be produced between clinicians and others as a result of research findings. Although the methodology for RCTs is generally well understood, many now recognize that they may be subject to more subtle problems of bias. The importance of a rigorous approach to the maintenance of blinding has been highlighted; failure to do so may result in bias and even the premature termination of a trial [13].

For practical reasons, the number of patients enrolled in controlled trials is usually limited to studies of efficacy, rather than the larger numbers of prolonged study that may be required to demonstrate safety. This potential problem has been addressed by meta-analyses, but these may be meta-analyses of the trial results themselves, rather than by analysis of the pooled original patient data from the trials concerned, with subsequent single analysis of the whole database. Although a substantial undertaking, this process has been recommended for improving the power of meta-analyses of RCTs [14].

In contrast, many observational studies simply observe the outcomes of patients who are given current therapy without any modification of either treatment, tests or any other interventions that may constitute an imposition for the patient. In this situation, written informed consent from the patient is frequently not necessary although ethical consideration of the project itself may still be necessary. It has to be said that the situation is variable between different countries and institutions and remains confusing. In general however, the administrative hoops to be jumped

through before setting up a patient database for an observational study are usually less.

Many institutions and departments of anaesthesia and critical care have developed their own substantive databases. Information technology has had a powerful effect on the feasibility of establishing such databases, which are primarily required for defining levels of activity, and for demonstrating the quality of healthcare. For example, anxieties over the quality of healthcare in cardiac surgery in the UK [15] led to the development of a quality of care database which has facilitated greater transparency than previously in terms of publishing clinical outcomes of cardiac surgery [16]. Quality and activity data are often requested now from healthcare purchasers to facilitate re-imburement. Finally, the use of suitable statistical techniques including propensity analysis and multivariate analysis with tests for interaction has stimulated many research groups to develop their own databases in order to conduct observational studies.

The practical advantages of observational databases are easily seen. Once set up, they can provide data on large numbers of patients in a short time frame compared to a RCT. Large numbers of data fields can and indeed should be collected although thought needs to be given as to how these are handled, and although the attention to data collection still needs to be meticulous this is often feasible within the organization of a clinical department or institution and should certainly be so within a research group. At a local or even national level, the costs in setting up an appropriately designed database for observational studies can be a fraction of those that would be encountered for a RCT. Also, if good follow-up procedures are in place, studies can run for a considerable period and provide data rarely feasible otherwise.

However, there are also drawbacks to observational databases which should limit their use. A major drawback is bias in treatment allocation, or lack of randomization. This can be offset by propensity analysis, a complex statistical technique that can reduce some of the potential for bias through lack of randomization by ensuring that potential confounders that are known are taken into account in the analysis [17,18]. This method is useful but is only as good as the confounding variables that have been gathered [19]. It can adjust for observed and known confounders, but only randomization will protect against bias from unknown confounders [9]. In addition, propensity scoring may include a number of irrelevant predictors, and inclusion of these in the model may reduce the effectiveness of the control on relevant predictors [20].

Whereas RCTs have pride of place in studies of efficacy, observational studies have been generally limited to safety or adverse event data analysis, particularly for drug trials. Safety studies in surgical

patients typically require a larger patient enrolment and for a longer duration than efficacy studies. However, since trials of efficacy are rarely adequately powered for adverse events, controlled trials that do identify concerns should be taken seriously. For example, a study of an i.v. coxib for pain relief following cardiac surgery suggested that, whilst there was no significant increase in individual adverse events in patients receiving active drug compared to placebo, there was an increase in adverse events overall [21]. A further study was required to be undertaken which confirmed these results [22], removing coronary artery bypass grafting (CABG) surgery from the product license. Problems in the use of coxibs in patients with cardiovascular disease are now well known.

In a recent review, Vandembrouke [23] has suggested that observational research should be restricted to questions that can meet the underlying assumption that the treatment is unrelated to the outcome being measured. This is easiest for studies on adverse effects, where the outcome under investigation is by definition unintentional if not actually unpredictable.

Let us consider two examples. First, consider a trial investigating the incidence of renal failure with a new opioid analgesic in patients undergoing hip replacement surgery. There is no known significant association between opioids and renal failure. Hip replacement surgery does not carry with it an innate risk of renal failure. The likely confounders are therefore probably related to the overall health status of the patients, since such patients may be elderly and might have pre-existing cardiovascular disease and/or renal dysfunction. The likely incidence of renal failure in this setting is very low, and an observational study with appropriate multivariate analysis with tests for interaction, or perhaps propensity analysis for relevant patient risk factors would seem appropriate.

Second, consider a trial investigating the incidence of renal failure with a drug with known adverse renal effects in patients undergoing CABG surgery. In this setting, the drug under investigation and the procedure have known adverse effects on renal function. The patient group is also at variably increased risk for renal dysfunction due to cardiovascular and secondary renal disease. The probable incidence of renal dysfunction is much higher than in the hip replacement study, due to the variable interaction of the effects of the drug, the clinical status of the patient and the nature of the surgery. In this setting an observational study will need to be much more sophisticated in its analysis and will need to take into account a much greater number of potential confounders. There is a risk that propensity analysis would not be adequate to eliminate confounders that otherwise would provoke the same outcome as the drug. In an ideal world, it could be argued that it would be better to carry out

a RCT in a pre-defined CABG population at specified risk and with renal failure as the primary outcome. Although this might be preferable, it might not be practical, and Hunter [19] has recently reviewed the reasons that RCTs may not be available for the assessment of adverse events associated with prescription drugs. These include inadequate power, inadequate study duration, monitoring limitations for adverse event detection in efficacy studies, early termination of an efficacy trial thus obscuring the true incidence and nature of adverse events and enrolment criteria that may actually exclude susceptible subgroups. This is particularly true since efficacy studies are often designed to try and exclude all possible confounders; thus the subjects included and randomized may be much younger and fitter than the likely patient cohort destined to receive treatment once a drug is licensed. This has been described as the difference between 'study patients' and those in the 'real world'.

RCTs are rarely undertaken comparing two proprietary medications due to the difficulties in organization and funding, and funding for adverse event studies is usually difficult to obtain from commercial sources once efficacy criteria have been established and a product licence obtained. Finally, once efficacy has been established, continued patient participation to obtain adverse event data may not be forthcoming, particularly if further invasive tests are involved. This problem has been highlighted most recently in the OPCAB vs. CABG studies [24]. How often do the results of observational studies and randomized trials disagree? Systematic reviews suggest that such disagreements are not uncommon [10,23]. When disagreements occur, it is usual that a promise of efficacy, suggested by observational studies is not borne out by RCTs. This again reinforces the observation that RCTs are undoubtedly superior to observational studies for evaluation of efficacy.

Differences in RCTs and observational studies for safety or adverse events are more complex. If the observational study has a role, it is surely in this area of clinical research, and recent commentators have not only reinforced this, but also highlighted important areas where the drawbacks of observational studies may be minimized both by study design and analysis [19,23,25].

In 1987 Royston and colleagues [26] first reported the use of aprotinin in cardiac surgery patients to minimize blood loss. Since that time, numerous studies and meta-analyses of RCTs have reinforced this finding. However, in January 2006, an observational study published by the Multicentre Study of Perioperative Ischemia (McSPI) Research Group in the *New England Journal of Medicine* concluded that the use of aprotinin in cardiac surgery patients was associated with a higher incidence of renal, cardiac and cerebral adverse events.

The authors of the manuscript were sufficiently convinced of the strength of their evidence to conclude that this drug should not be used in patients undergoing cardiac surgery [12]. A smaller observational study published recently has suggested an increased incidence of renal dysfunction with aprotinin [27].

Needless to say, this issue is one of enormous concern to cardiac anaesthesiologists, cardiac surgeons and patients. Bleeding is one of the largest contributors to risk of death after heart surgery. Since the blood-sparing effect of aprotinin was first reported, research has repeatedly shown that the drug reduces blood loss and the need for transfusion [28–31] both of which are associated with adverse outcomes in cardiac surgery. Thus we face the possibility that either the treatment, or the lack of treatment, could be associated with significant adverse outcomes.

Correspondence about the article has not surprisingly highlighted the differences between this study and the published RCTs, and has questioned the study design, the ability of propensity analysis to adequately compensate for the lack of randomization, suggested inconsistencies between this study and others published by the same group and a number of other issues which are contained within the relevant correspondence [12,32]. Both the original article and the resulting correspondence are well worth reading.

In this issue, Dr Royston asks the question Aprotinin; Friend or Foe, and delivers a critique of the study published in the *New England Journal*. Dr Royston was first author on the first report of the blood-sparing effect of aprotinin, and has been involved in a number of clinical trials concerning its use. His article published here gives another important perspective to the use of aprotinin in cardiac surgery, and one that is at odds with the study of Mangano and colleagues [12].

Whilst the experts continue to disagree, what should the rest of us do? The answer is ... tread very carefully! The US Food and Drug Administration (FDA), and three specialist societies in cardiac anaesthesia (Society of Cardiovascular Anaesthesiologists [33], European Association of Cardiothoracic Anaesthesiologists [34], Association of Cardiothoracic Anaesthetists [35]) are all unanimous in suggesting that, at this time and where the drug is available, the use of aprotinin should be guided by carefully weighing the benefits and drawbacks of its use in each patient.

However, the most recent review of the evidence by the FDA has highlighted the need for ensuring that data sources for all observational studies are made available wherever possible. This is surely a duty both for independent researchers and for the pharmaceutical industry [36,37]. Failure to do so would mean that everyone, regulators and clinicians, would be working in the dark whilst dealing with serious issues about

drug safety. It is difficult to see how this can benefit patients.

At least one major study that is currently underway may serve the shed further light on this issue, but we have to consider the possibility that it may not, and that the quality of evidence we have on this subject may not improve for some considerable time, if at all. In which case clinicians will need to take account of the published data available, decide which of the studies are best designed to answer those questions they consider pertinent and come to a conclusion based not only on the published evidence but also on their own clinical experience and judgement. This dilemma serves to underline the importance that postgraduate medical education should continue to attach to teaching an understanding of the research process and to interpreting its results.

Conflict of interest statement

Dr Feneck is a member of the board of directors of EACTA, an organization which has been in receipt of a number of unrestricted educational grants from Bayer, manufacturer of Trasyolol. He is also a member of the McSPI research group.

R. O. Feneck

References

1. Jackson RG, Stamford JA, Strunin L. The canary is dead. *Anaesthesia* 2003; 58: 911–912.
2. Lee KP, Boyd EA, Holroyd-Leduc JM, Bacchetti P, Bero LA. Predictors of publication: characteristics of submitted manuscripts associated with acceptance at major biomedical journals. *Med J Aust* 2006; 184: 621–626.
3. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002; 359: 614–618.
4. Fisher RA. The design of experiments, 8th edn, 1971. In: Bennet JH, ed. *Statistical Methods, Experimental Design, and Scientific Inference*. Oxford: Oxford University Press, 1995.
5. Chalmers I. Why transition from alternation to randomisation in clinical trials was made. *BMJ* 1999; 319: 1372.
6. Day SJ, Altman DG. Statistical notes; blinding in clinical trials and other studies. *BMJ* 2000; 321: 504.
7. House of Lords Science and Technology (6th Report) on Complementary and Alternative Medicine. November 2000. www.publications.parliament.uk
8. Tobin JR, Shafer SL, Davis PJ. Pediatric research and scholarship: another Gordian Knot? *Anesth Analg* 2006; 103: 43–48.
9. Deeks JJ, Dinnes J, D'Amico R *et al.* International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003; 7: 1–173.
10. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998; 317: 1185–1190.

11. Julian DG. What is right and what is wrong about evidence-based medicine? *J Cardiovasc Electrophysiol* 2003; 14(9 Suppl): S2–S5.
12. Mangano DT, Tudor IC, Dietzel C. Multicentre Study of Perioperative Ischemia Research Group, Ischemia Research and Education Foundation. The risk associated with aprotinin in cardiac surgery. *New Engl J Med* 2006; 354: 353–365.
13. Kirk-Smith MD, Stretch DD. Evidence-based medicine and randomized double-blind clinical trials: a study of flawed implementation. *J Eval Clin Pract* 2001; 7: 119–123.
14. Clarke MJ, Stewart LA. Systematic reviews of randomized controlled trials: the need for complete data. *J Eval Clin Pract* 1995; 1: 119–126.
15. The Report of the Public Inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–1995; Learning from Bristol. July 2001. www.bristol-inquiry.org.uk/final_report
16. www.heartsurgery.healthcarecommission.org.uk
17. Blackstone EH. Comparing apples and oranges. *J Thorac Cardiovasc Surg* 2002; 123: 8–15.
18. D'Agostino Jr RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; 17: 2265–2281.
19. Hunter D. First, gather the data. *New Engl J Med* 2006; 354: 329–331.
20. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; 127: 757–763.
21. Ott E, Nussmeier NA, Duke PC *et al.* Multicenter Study of Perioperative Ischemia (McSPI) Research Group; Ischemia Research and Education Foundation (IREF) Investigators. Efficacy and safety of the cyclooxygenase 2 inhibitors parecoxib and valdecoxib in patients undergoing coronary artery bypass surgery. *J Thorac Cardiovasc Surg* 2003; 125: 1481–1492.
22. Nussmeier NA, Whelton AA, Brown MT *et al.* Complications of the COX-2 inhibitors parecoxib and valdecoxib after cardiac surgery. *New Engl J Med* 2005; 352: 1081–1091.
23. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004; 363: 1728–1731.
24. Feneck R. OPCAB surgery: time for a reappraisal? *J Cardiothorac Vasc Anesth* 2004; 18: 253–255.
25. Etminan M, Samii Etminan M, Samii A. Pharmacoepidemiology I: a review of pharmacoepidemiologic study designs. *Pharmacotherapy* 2004; 24: 964–969.
26. Royston D, Bidstrup BP, Taylor KM, Sapsford RN. Effect of aprotinin on need for blood transfusion after repeat open-heart surgery. *Lancet* 1987; 2: 1289–1291.
27. Karkouti K, Beattie WS, Dattilo KM *et al.* A propensity score case-control comparison of aprotinin and tranexamic acid in high-transfusion-risk cardiac surgery. *Transfusion* 2006; 46: 327–338.
28. Young C, Horton R. Putting clinical trials into context. *Lancet* 2005; 366: 107–108.
29. Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials* 2005; 2: 218–229.
30. Sedrakyan A, Treasure T, Eleftheriades JA. Effect of aprotinin on clinical outcomes in coronary artery bypass graft surgery: a systematic review and meta-analysis of randomized clinical trials. *J Thorac Cardiovasc Surg* 2004; 128: 442–448.
31. Henry DA, Moxey AJ, Carless PA *et al.* Anti-fibrinolytic use for minimising perioperative allogeneic blood transfusion. *Cochrane Database Syst Rev* 2001; 1: CD001886.
32. Sedrakyan A, Atkins D, Treasure T. The risk of aprotinin: a conflict of evidence. *Lancet* 2006; 367: 1376–1377.
33. www.scahq.org
34. www.eacta.org
35. www.acta.org.uk
36. Avorn J. Dangerous deception – hiding the evidence of adverse drug effects. *New Engl J Med* 2006; 355: 2169–2171.
37. Hiatt WR. Observational studies of drug safety – aprotinin and the absence of transparency. *New Engl J Med* 2006; 355: 2171–2173.