ANZJSurg.com

How to dissect surgical journals: XIV – Clinical trials*

The necessary accomplishments of the carpenter are that his work not be warped and that joints are aligned. The work must be smoothed out with a plane and not polished in a haphazard way to disguise defects. It is important that the work does not warp afterward.

Miyamoto Musashi (Samurai Warrior, 17th Century)

Performing a clinical trial is like running a business. Investigators have to find the capital, outline a clear plan of action, keep everyone happy, maintain good lines of communication, ensure nobody is taking shortcuts, complete tasks in a timely manner and, at all times, be willing to compromise perfection for what is achievable. A key aspect of many trials is the ability to recruit a disjointed, sometimes indifferent and often suspicious group of hospital staff as collaborators. It is far easier to comment on published trials than it is to perform them.

Clinical trials are prospective studies that evaluate an intervention, usually a drug or a procedure, on health outcomes after the patients have been randomized into groups

Three-quarters of surgical trials evaluate drugs such as antimicrobial agents, anticoagulants, analgesics and anticancer agents.[1] Probably because they are easier to get funded and not as complex as trials involving operations or other aspects of surgical care. Surgical trials are usually phase III or phase IV studies.

Clinical trials that evaluate drugs are divided into four phases:

- Phase I: Initial safety trial in normal volunteers.
- Phase II: Pilot trial to evaluate efficacy in selected populations of patients.
- Phase III: Accrual of additional safety and efficacy prior to regulatory approval.
- Phase IV: Post-approval clinical trials.

Investigators sometimes use the word 'pilot' as an excuse for publishing a small study. It avoids the rigours of registering the trial and declaring formal estimates of the required sample size. Nevertheless, pilot studies can provide useful information. Chemla and Morsy[2] compared two vascular grafts. The graft under evaluation, a decellularized bovine ureteric graft, had already been evaluated in 'several non-randomized series' (phase II trials). They performed a 'prospective, pilot randomized study' in 60 patients and found that the new graft had a clinically unacceptable patency rate at 1 year – 28% versus 48% – even though the differences were not statistically significant. The outcome of the study was withdrawal of the new graft from the market with continued research aimed at producing a less thrombogenic second-generation graft. Under these circumstances, it seems reasonable to go with the trend.

Many journals insist that clinical trials are enrolled in a public registry.[3] These registries provide a unique trial number and access to a comprehensive 'Minimal Registration Data Set' (http://register.clinicaltrials.gov). The registration of clinical trials at their initiation allows others to view the protocol and identify changes that are made during or after completion of the trial – such as altering the study aim based on the results and revising the calculations that were made to determine the sample size.

Consort

The Consolidated Standards of Reporting Trials (CONSORT) statement is an essential reading. Ignore the brief paper-based version that was published simultaneously in several medical journals. Instead, go to their website (http://www.consort-statement.org) and click onto 'Resources' and then 'Downloads'. The discussion document includes a 22-item checklist with linked text and diagrams. In the midst of these explanations are 12 core issues:

- Selection: A transparent account of the criteria for admission of participants into the trial.
- Intervention: A description of the intervention that would allow other investigators to accurately replicate the study.
- Aim: A clearly stated aim within the Introduction.
- Outcome events: The declaration and unambiguous definition of the main outcome events.
- Sample size: A formal and prospective determination of sample size.
- Randomization: A valid process to generate the random allocation sequence.
- Concealment: How the randomization details were concealed from persons entering patients into the trial.
- Blindness: A statement that assessors of the primary outcome events were not aware of the group allocation for each patient.
- Flow of patients: An account of the participants who were excluded from study or withdrawn from study.
- Baseline comparison: Declaration of factors that are descriptive of the participants and risk factors that might confound the interpretation of the results.
- Exclusions from analysis: Comparisons should usually be based on an 'intention to treat'.

*The 16 articles in this series are being made available on ANZJSurg.com as an eBook.

© 2011 The Author
ANZ Journal of Surgery © 2011 Royal Australasian College of Surgeons

- Comparison of outcomes: A statement about the magnitude and reliability of the observed differences between the groups beyond the display of '*P*-values', that is, use of indices of risk and confidence intervals.

These criteria have been evaluated in surveys of surgical trials.[4–8] Although there has been a slow but progressive improvement in the reporting of methodological standards in surgical trials, there are still some special problems.

The need for an overview

A particularly successful requirement of the CONSORT statement has been the need for authors to include a diagram of the flow of patients. It draws attention to the number of patients who were withdrawn from study or lost to follow-up. In the absence of such a diagram, it is often difficult to confirm that all randomized patients were included in the final analyses or, if they failed to undergo surgery, were accounted for as withdrawals from study.

It is also useful to have access to an outline of the architecture of the study. These diagrams are mandatory in protocols and grant applications, but usually fail to make it into the final article. It is often useful for readers to sketch an outline that includes time lines and the complexities of trial closeout. It makes the text more transparent.

Factorial designs introduce an interesting variation in the design of trials. They allow for the evaluation of two interventions, and their interactions, without great increases in complexity (the downside is the need to recruit more patients). Dunkelgrun *et al*.[9] used such a 2 × 2 factorial design (Latin square) to evaluate beta-blocker and statin therapies to reduce perioperative cardiac events. They randomized the patients into four groups:

Group 1: Beta-blocker	Group 2: No Intervention (control)
Group 3: Beta-blocker + Statin	Group 4: Statin

Nelson *et al*.[10] used a more complex 2 × 2 × 2 factorial design to evaluate the healing of venous ulcers. Patients were first randomized to receive either pentoxifylline or placebo, then they were randomized to receive either knitted viscose or hydrocolloid dressings and, finally, they were randomized to receive either single-layer or four-layer bandages. This provided an evaluation of the 'relationship between drugs, dressings, and bandages in venous ulcer healing'.

Standardizing operative interventions

Trials involving operation require an explicit account of the intervention so that it can be reproduced by other surgeons Video files are helpful and they can often be submitted as supplementary material and made available on the journal web site. However, about one-quarter of the trials evaluating operations fail to provide an adequate account of operative technique. For example, the German Angioplasty Bypass Surgery Investigation[11] described the method of coronary artery bypass grafting (CABG) with the solitary comment that there 'were no requirements concerning technical details of the interventions'. Greater detail would have been helpful as they concluded that 'patients treated with CABG were more likely to sustain an acute myocardial infarction at the time of the procedure'.

Heute Billroth, morgen tod
(Today Billroth, tomorrow death)

Anonymous

The relative skills of surgeons may be a confounding influence in clinical trials. Lawrence[12] commented that 'The relative risk of a bad outcome from one technique versus another may be far less than that between an inexperienced and an experienced surgeon'. This is especially so when there is a learning curve associated with the adoption of a new technique. Bonenkamp *et al*.[13] provided a positive example when they evaluated ablative surgery for cancer of the stomach and reported that 'surgeons were given a videotape and an instruction booklet, and regular meetings for technical back-up were held'.

Less than one-in-five surgical trials monitor compliance with the surgical protocol. Surgeons, like other clinicians, are highly motivated towards providing optimal care for each patient. Although laudable in all other circumstances, it can lead surgeons to deviate from a protocol in the belief that they are doing the best thing for their patient. This is most likely to occur early in a trial if all of the participating surgeons have not fully accepted the appropriateness of each arm of the study or near the conclusion of the trial if surgeons perceive that one group of patients is either doing better or doing worse than the others.

Concealing the intervention

It is important that patients, clinical staff, data collectors and those analysing data are not aware of the allocation. The term 'allocation concealment' has been applied to the general process of keeping the randomization process confidential. It includes:

- *Masking*: the technique used to hide the randomization schedule from those assigning participants into groups, for example, opaque envelopes, use of a randomization centre.
- *Blinding*: the overall process of non-disclosure of the allocated group after randomization, for example, matching tablets, standard wound dressings, independent assessors.

Although masking is always appropriate, blinding can be difficult to achieve. You can usually assume that single-blind means that the patients were unaware of their allocated group. However, the statement that a trial is 'double blind' is ambiguous. The second 'blind' could apply to the attending surgical team, the persons collecting the data or the persons analysing the data.

Sometimes authors confuse the randomization technique with the masking process. Kotsch *et al*.[14] studied the effect of prednisolone on inflammation in donor livers and stated that: 'Randomization was performed with sealed envelopes'. Sealing envelopes is not a randomization technique.

© 2011 The Author
ANZ Journal of Surgery © 2011 Royal Australasian College of Surgeons

Investigators should assume that everyone participating in a clinical trial will 'cheat' at every opportunity

It is important to anticipate that vials will be opened and tasted and that envelopes will either be opened and 'lost' or held up to light in order to view their contents. If envelopes are to be used, then they must be sealed, opaque and serially numbered. Investigators can minimize 'cheating' via personal detailing, education and vigilant monitoring. But sometimes, the consequences of administering an agent cannot be hidden from the attending medical staff. Boffard *et al*.[15] evaluated the use of recombinant activated factor VIIa (rFVIIa) in trauma. They 'blinded' patients, clinicians and study personnel. However, there was no concealment of laboratory tests and rFVIIa effects the coagulation profile.

Investigators sometimes blind the surgical procedure. Hall *et al*.[16] evaluated infants after open and laparoscopic pyloromyotomy for pyloric stenosis. They wanted to blind the nature of the procedure, and to achieve this, 'three opaque dressings were applied at the end of the surgical procedure to all infants, regardless of the treatment received. Dressings were removed for wound inspection after taking the decision to discharge the infant from hospital'. They observed a significant reduction in analgesia requirement after the laparoscopic procedure. More than a decade earlier, Johnson and Majeed[17] compared laparoscopic with small-incision cholecystectomy and 'Identical wound dressings were applied in both groups so that carers could be kept blind to the type of operation'. They found that 'Laparoscopic cholecystectomy takes longer to do than small-incision cholecystectomy and does not have any significant advantages in terms of hospital stay or postoperative recovery'. It is interesting that this well-conducted trial in a premier journal had little impact on the growth of laparoscopic surgery.

Interim analyses

Interim analyses refer to the repeated analyses of data as they accumulate. They are either formally planned or they are done in an uncontrolled fashion, which increases the chance of spurious claims of statistical significance.

Statistical techniques have been described for dealing with planned interim analyses. As the name vaguely suggests, the DISPACT trial was designed to compare two different techniques of distal pancreatectomy.[18] The investigators aimed to recruit about 336 patients from up to 20 centres. They reasonably commented that 'Due to uncertainties about the effect sizes of the primary endpoint, an *a priori* planned interim analysis of the primary endpoint will be conducted after 224 evaluable patients are selected in order to reassess the initially planned sample size. Since pre-existing evidence was limited, our initial sample size calculation is based on uncertain assumptions and may need to be modified in a planned interim analysis'.

In a sequential trial, the decision to continue further collection of data depends on the information that has already been collected. The sample size depends upon the results that are obtained during the course of the study (i.e. there are data-dependent stopping rules). This is attractive because there are financial and ethical reasons why it is undesirable to continue a study once a reliable conclusion has been reached. The trial ceases if an interval analysis reveals a statistically significant difference between the groups; otherwise, it continues until the sample size is reached.

Sequential trials require:
A single study objective
A short treatment response
Reliable monitoring techniques

In 1965, Joyce and Welldon[19] performed a clinical trial to evaluate 'the objective efficacy of prayer'. They quoted Francis Galton (1883) who asked, 'Do sick persons who pray or are prayed for recover on the average more rapidly than others?' If you would like an interesting exercise, review their publication using the CONSORT criteria.

Adverse events and follow-up

It is crucial that surgical trials measure harm as well as benefit. The results of surgery are often a trade-off between short-term risks and long-term effects. Up to one-third of surgical trials fail to give an adequate account of adverse events. It is difficult to estimate the value of interventions in the absence of this information. Of particular concern is the failure of reports of major elective operations to report the numbers of deaths (absence of comment cannot be interpreted as an absence of deaths).

Ethical issues can also lead to the premature closure of trials. Morak *et al*.[20] evaluated the effects of adjuvant intra-arterial chemotherapy and radiotherapy versus surgery alone in resectable pancreatic and periampullary cancer. In their words: 'several important clinical trials demonstrated a survival benefit of adjuvant chemotherapy after curative surgery, which was not yet reported when this study was initiated. Because our control group (observation group) did not receive any form of adjuvant treatment, it was considered unethical to continue to include patients'. They studied 120 patients, rather than the anticipated 240, and reported a reduction in the development of liver metastases and a prolonged time to progression of the cancer.

Multicentre trials

Multicentre trials use the same protocol simultaneously for patients at different sites. They require careful administration, good communications and a close monitoring of compliance and quality control. Large multicentre trials are major undertakings with independent committees for monitoring compliance and analysing data. Another positive attribute is their formal approach to pilot studies.

Pilot studies
Estimate accrual rates
Determine suitable end points
Allow estimation of the required sample size
Evaluate measures of compliance

© 2011 The Author
ANZ Journal of Surgery © 2011 Royal Australasian College of Surgeons

Analyse the learning curve of participants.
Design forms
Determine logistical issues.

Reports by multicentre studies should discuss the possibility of variations between the contributing centres. Some multicentre trials stratify by the centre, which is a very appropriate approach as it allows for better evaluations when appreciable differences are observed between the centres and decisions need to be made about contentious issues such as centre-adjusted analyses. Note that the outcome of some procedures is more influenced by the nature of the hospital than specific surgeons. Both surgeons and institutions can be risk factors.

Cluster trials

Cluster trials study groups of patients rather than individuals. They are appropriate when an intervention can only be effectively implemented, and isolated from the control group, by introducing it in a contained environment. For example, Doig *et al*.[21] recruited patients from 27 intensive care units to evaluate the effect of evidence-based feeding guidelines compared with routine nutritional care. The staff at participating units were actively indoctrinated about the evidence-based guidelines and received feedback from audits. It can be impossible to quarantine this type of effort. The planning and analysis of cluster trials is complex – see the CONSORT extension statement for details.

Equivalence trials

Equivalence trials seek to demonstrate that a new intervention is as good as the established intervention. The reasons for performing this type of trial can relate to reduced costs, less side effects or greater ease of performance. The sample size requirements for equivalence trials are onerous and special care must be taken to reduce the probability of a type II (false-negative) error to an acceptable level.

Equivalence trials usually relate to drugs, but they can be applied to procedures. Kapur *et al*.[22] designed a multicentre equivalence trial to evaluate coronary artery procedures in diabetics. Their aim was to address the hypothesis that optimal percutaneous coronary intervention is not inferior to modern CABG as a revascularization strategy for diabetics with multivessel or complex single-vessel coronary disease.

Be aware that some authors re-label trials with negative results as equivalence trials in an effort to enhance their appeal. They are inevitably underpowered and carry an unacceptable risk of a type II error. Genuine equivalence trials are designed to avoid that problem, but the issues are complex and extend beyond the mere consideration of sample size – see the CONSORT extension statement on non-inferiority and equivalence trials for details.

The limitations of clinical trials

A number of operations that were introduced with great optimism were subsequently rejected after the performance of a clinical trial. Examples include internal mammary ligation for angina pectoris and extracranial–intracranial arterial bypass to reduce the risk of ischaemic stroke. Such historical precedents – as well as the chaotic introduction of laparoscopic cholecystectomy in the 1980s – emphasize the need for trials to define the role of new interventions.

Clinical trials cannot resolve every issue of concern to surgeons. Sometimes, the logistic problems are too formidable; modifications of operations can evolve rapidly and be a 'moving target'; on other occasions, there are ethical matters that cannot be settled; and, all too often, willing investigators may lack the necessary resources.

Efficacy is the extent that an intervention can produce a beneficial outcome *under ideal circumstances*.
Effectiveness is the extent that an intervention can produce a beneficial outcome *in routine clinical practice*.

Although they produce high-quality evidence, clinical trials can be remote from everyday clinical practice. They are also limited in their ability to test for safety, because they only evaluate relatively small numbers of patients and they tend to exclude patients who are susceptible to complications (selection bias).

Key points

- Clinical trials are prospective studies that evaluate an intervention, usually a drug or a procedure, after the patients have been randomized into groups.
- Three-quarters of surgical trials evaluate drugs such as antimicrobial agents, anticoagulants, analgesics and anticancer agents.
- Clinical trials should be enrolled in a public registry at their initiation so that readers can detect changes made during or after completion of the trial.
- The CONSORT statement is an essential reading. The discussion document includes a 22-item checklist with linked discussions and diagrams.
- A particularly successful requirement of the CONSORT statement has been the need for authors to include a diagram of the flow of patients.
- Evaluation of operations raises three issues. First, the standardization of operative techniques. Second, the skill of the surgeons participating in the study. And, finally, compliance with the study protocol.
- Masking is the technique used to hide the randomization schedule from those assigning participants into groups, for example, opaque envelopes.
- Blinding refers to the overall process of non-disclosure of the allocated group.
- Statistical techniques have been described for dealing with interim analyses, but they should be planned in advance.
- Patients need to be followed up for as long as it takes to measure harm as well as benefit.
- Multicentre studies should discuss the variations between the contributing centres.
- In cluster trials, the basic unit is groups of patients rather than individuals.

© 2011 The Author
ANZ Journal of Surgery © 2011 Royal Australasian College of Surgeons

- Equivalence trials seek to demonstrate that a new intervention is as good as the established intervention.
- There are a limited number of clinical situations that can be evaluated in clinical trials.

References

1. Hall JC, Mills B, Nuygen H, Hall JL. Methodological standards in surgical trials. *Surgery.* 1996; **119**: 466–72.
2. Chemla ES, Morsy M. Randomized clinical trial comparing decellularized bovine ureter with expanded polytetrafluoroethylene for vascular access. *Br. J. Surg.* 2009; **96**: 34–9.
3. Haug C, Gotzsche PC, Schroeder TV. Registries and registration of clinical trials. *N. Engl. J. Med.* 2005; **353**: 2811–2.
4. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized clinical trials in surgery. *Surgery.* 1994; **115**: 707–12.
5. Hall JC, Platell C, Hall JL. Surgery on trial: an account of clinical trials evaluating operations. *Surgery.* 1998; **124**: 22–7.
6. McLeod RS. Issues in surgical randomized controlled trials. *World J. Surg.* 1999; **23**: 1210–4.
7. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ.* 2002; **324**: 1448–51.
8. Ellis C, Hall JL, Khalil AA, Hall JC. The evolution of methodological standards in surgical trials. *ANZ J. Surg.* 2005; **75**: 874–7.
9. Dunkelgrun M, Boersma E, Schouten O *et al*., for The Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group. Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing noncardiovascular surgery: a randomized controlled trial (DECREASE-IV). *Ann. Surg.* 2009; **249**: 921–6.
10. Nelson EA, Prescott RJ, Harper DR, Gibson B, Brown D, Vaugham Ruckley C. A factorial, randomized trial of pentoxifylline or placebo, four-layer or single-layer compression, and knitted viscose, or hydrocolloid dressings for venous ulcers. *J. Vasc. Surg.* 2007; **45**: 134–41.
11. Hamm CW, Reimers J, Ischinger T *et al*. A randomized study of coronary angioplasty compared with bypass surgery in patients with symptomatic multivessel coronary disease. *N. Engl. J. Med.* 1994; **331**: 1037–43.
12. Lawrence W Jr. Some problems with clinical trials. *Arch. Surg.* 1991; **126**: 370–8.
13. Bonenkamp JJ, Songun I, Hermans J *et al*. Randomised comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. *Lancet.* 1995; **345**: 745–8.
14. Kotsch K, Ulrich F, Reutzel-Selke A *et al*. Methylprednisolone therapy in deceased donors reduces inflammation in the donor liver and improves outcome after liver transplantation: a prospective randomized controlled trial. *Ann. Surg.* 2008; **248**: 1042–50.
15. Boffard DK, Riou B, Warren B *et al*., for the NovoSeven Trauma Study Group. Recombinant factor VIIa adjuvant therapy for bleeding control in severely injured trauma patients: two parallel randomized, placebo-controlled, double-blind clinical trials. *J. Trauma.* 2005; **59**: 8–18.
16. Hall NJ, Pacilli M, Eaton S *et al*. Recovery after open versus laparoscopic pyloromyotomy for pyloric stenosis: a double-blind multicentre randomised controlled trial. *Lancet.* 2009; **373**: 390–8.
17. Johnson AG, Majeed AW. Randomized clinical trial of conventional cholecystectomy versus minicholecystectomy. *Br. J. Surg.* 1998; **85**: 1017–8.
18. Diener MK, Knaebel HP, Witte ST *et al*., DISPACT Trial Group. DISPACT trial: a randomized controlled trial to compare two different surgical techniques of DIStal PAnCreaTectomy – study rationale and design. *Clin. Trials.* 2008; **5**: 534–45.
19. Joyce CRB, Welldon RMC. The objective efficacy of prayer: a double-blind clinical trial. *J. Chronic Dis.* 1965; **18**: 367–77.
20. Morak MJM, van der Gaast A, Incrocci L *et al*. Adjuvant intra-arterial chemotherapy and radiotherapy versus surgery alone in resectable pancreatic and periampullary cancer: a prospective randomized controlled trial. *Ann. Surg.* 2008; **248**: 1031–41.
21. Doig GS, Simpson F, Finfer S *et al*. Effect of evidence-based feeding guidelines on mortality of critically ill adults: a cluster randomized controlled trial. *JAMA.* 2008; **300**: 2731–41.
22. Kapur A, Malik IS, Bagger JP *et al*. The Coronary Artery Revascularisation in Diabetes (CARDia) trial: background, aims, and design. *Am. Heart J.* 2005; **149**: 13–9.

John C. Hall, MS, DS, FRACS
Surgery (Royal Perth Hospital),
The University of Western Australia,
Perth, Western Australia, Australia

doi: 10.1111/j.1445-2197.2010.05648.x

© 2011 The Author
ANZ Journal of Surgery © 2011 Royal Australasian College of Surgeons