ANZJSurg.com

How to dissect surgical journals: V – Patients*

Law Number Four: The patient is the one with the disease.
Samuel Shem (The House of God)

Case studies are focused and direct – they may amuse, warn, or inform. It is hard for their construct to mislead. On the other hand, you need to be on your guard when authors make recommendations after studying groups of patients. Generalisations and pitfalls abound.

Selection biases

Patients entered into a study at 'The Megatropolis Centre of Clinical Excellence' will differ from those at 'The Shady Oaks District Hospital'. They will have passed through a number of 'selection pressures'. Ambulances and retrieval teams bypass smaller centres when they have a critically ill patient. Smaller hospitals send their complex patients to central hospitals. Some patients seek 'the best'.

> You cannot assume that patients who are written about in surgical journals are similar to the patients in your surgical practice.

A disproportionate number of publications come from highly specialised units. The staff have narrow skills and they have access to advanced technology. Smaller hospitals usually lack the infrastructure needed to support large clinical studies. This means that the surgical literature is biased towards the management of 'difficult' patients. This is compounded if the investigators just ask colleagues to nominate patients (convenience sampling) – it can bias towards the unusual.

If you wanted to survey the views of users of the hospital emergency department, you could recruit respondents by advertising in the local newspaper. However, the sample would be skewed in favour of users who liked to read community newspapers and were highly motivated to respond. It might favour those with a grudge and ignore the 'silent majority'. Asking a random sample of patients attending the clinic on a particular day would avoid this problem. But this approach might still result in a non-representative sample. Patients who attend on a Monday morning in the middle of winter will have a different profile from patients who attend on a Saturday night in the middle of summer. The latter is least likely because investigators might lack the funds to pay staff penalty rates to recruit patients outside office hours. The point is that selection bias is an issue even in seemingly simple studies.

Entry criteria should be clearly stated and unambiguous. But accurately defining the condition under evaluation can be difficult. Mishra *et al*.[1] performed a systematic review of prostatic massage for chronic prostatitis. They commented that chronic prostatitis, sometimes called the chronic pelvic pain syndrome, is characterised by 'pain in the perineum, pelvis, suprapubic area, or the external genitalia, with a variable degree of voiding or ejaculatory disturbance'. They noted the existence of a 'myriad of etiopathologic mechanisms for chronic prostatitis – some not even involving the prostate gland'. They only found one admissible clinical trial and it was of low quality. They also noted that 'the design of a sham procedure to mimic prostatic massage may be extremely difficult'. It is impossible to resolve issues when both the disease and the intervention are vague.

Exclusion criteria are the situations that will prevent patients, who otherwise fulfil the entry criteria, from entering the study. Exclusion criteria usually relate to patient safety, the extent of the primary condition, the extent of co-morbidity (coexisting illness), ethical issues (particularly the inability to provide informed consent because of illiteracy or poor mentation), and a high likelihood of poor compliance with the study protocol.

> Diagnostic Purity Bias: The exclusion of patients with comorbidities leads to the accumulation of an unrepresentative group of patients

It is informative when investigators make relevant comments. For instance, when reporting on a trial evaluating arthroscopic surgery for osteoarthritis of the knee, Kirkley *et al*.[2] commented that: 'Our study had limitations . . . only 68% of the patients who were evaluated for participation were deemed eligible and ultimately assigned to treatment. However, the majority of the excluded patients had substantial malalignment (38%) or declined participation (35%). The stringent inclusion and exclusion criteria ensured appropriate patient selection and optimised the chance of identifying a benefit of surgery. Thus, we do not believe that the participants in this trial were less likely to have a response to arthroscopic therapy than those treated in the community.' Such comments give readers the opportunity to draw their own conclusions, which is always a good thing. It would be interesting to know why 35% of the patients declined to enter the study. It is always difficult for patients to maintain an

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

© 2010 The Author
Journal compilation © 2010 Royal Australasian College of Surgeons

equipoise when an operation is being compared with a 'no intervention' control groups. The default is to avoid invasive therapy.

Lasagna's Law

Louis Lasagna from Tufts University observed that investigators are too optimistic about the numbers of patients that will enter into a study. There are two main reasons for this. First, investigators often fail to perform pilot studies that accurately documents the availability of patients. It is no good just relying on coded admission rates – coding systems in hospitals are notorious for their error rates and many will not be suitable for study. Second, investigators usually under-estimate the extent of 'wastage' – patients who could have entered the study may fail to do so because of logistical problems. Doctors who are not investigators are reluctant to enter patients; in fact, they often regard it as a nuisance.

> Lasagna's Law: Availability of suitable patients falls dramatically when the study starts and rapidly rises on completion of the study.

Lindström *et al*.[3] evaluated whether smoking cessation would reduce the incidence of post-operative complications after primary hernia repair, laparoscopic cholecystectomy, and hip or knee arthroplasty. They calculated a need for 586 patients, but only recruited 117. The causes were a reduction in the numbers of smokers, the link between smoking and alcohol abuse, a higher proportion of smoking among immigrants (the study was conducted in Sweden), and a reluctance to stop smoking.

The control group

Results are always improved by omitting controls. In 1958, Mitchell *et al*.[4] suggested that bilateral ligation of the internal mammary arteries benefited patients with angina by increasing flow in the coronary arteries via collaterals. It was enthusiastically reported in the popular press, including *Reader's Digest*. The issue was resolved when Cobb *et al*.[5] performed a double-blind study using a control procedure consisting of parasternal skin incisions. The groups had similar outcomes. One of the patients 'who had been unable to work because of his heart disease, was almost immediately rehabilitated and was able to return to his former occupation. He reported a 100 per cent improvement at six months and 75 per cent improvement after a year. His arteries were not ligated'. It is of interest that about 10% of trials evaluating operations compared a procedure with either a non-operative intervention or no intervention.[6] These studies have currency for conditions that have a natural tendency to improve: they allow for the natural history of the condition.

The surgical literature contains too many uncontrolled studies. Centa *et al*.[7] evaluated 29 octogenarians undergoing laparoscopic abdominal aortic aneurysm (AAA) repair. They concluded that this technique is 'complementary' to open and endovascular aneurysm repair, and that these results 'encourage us to offer laparoscopic AAA repair in good surgical risk octogenarians'. All this from studying a single relatively small group of patients. At best, this report support the further evaluation of laparoscopic AAA repair in controlled studies. Although such reports flourish within the surgical literature, the adoption of interventions based on such flimsy evidence is a cause of concern.

> Good ideas do not always work in practice

Clinical trials usually include a control group that receives the current 'gold-standard' therapy. For example, it would be ethically and scientifically wrong to include a 'no treatment' group in a study evaluating the ability of a new antibiotic to prevent sepsis after colorectal surgery. Once an effective treatment has become established, the essential issue is whether the new treatment is superior to the best current therapy.

The interpretation of clinical trials can be confounded by the 'placebo effect'. Placebo comes from the Latin for 'I shall please'. It was used in medieval times for the sham mourners who were hired to wail at funerals. The placebo effect was originally described in the 1920s when the Western Electric Company was investigating the effect of illumination on the typing efficiency of secretaries at its Hawthorne plant in Chicago. They noted a rise in productivity that was independent of the degree of illumination. The mere fact that a study was being performed was enough to raise productivity.

A placebo is an inactive pill, liquid, or powder that has no intrinsic treatment value. An example of the placebo effect is the relief of pain after an injection of saline. Placebos can be associated with both positive and negative effects. The term 'nocebo', Latin for 'I will harm', is sometimes used to describe adverse effects associated with a placebo.

Look carefully at studies that include 'standard care' as the control group. First, what is standard care? In some healthcare systems 'standard care' is a euphemism for inattentive or poor care. Second, has allowance been made for a placebo effect?

> Always question the validity of the control group

Kshettry *et al*.[8] randomised patients undergoing cardiac surgery to receive either 'complementary therapy (preoperative guided imagery training with gentle touch or light massage and postoperative music with gentle touch or light massage and guided imagery) or standard care'. It can be argued that any positive intervention that engages the patient may be beneficial for pain control and patients' perceptions of wellbeing i.e., it has a 'placebo effect'. In other words, the effect is not specific to that intervention. It could be achieved in other ways. Nevertheless, a juxtaposed invited commentary failed to make that distinction and stated that 'Partnership

© 2010 The Author
Journal compilation © 2010 Royal Australasian College of Surgeons

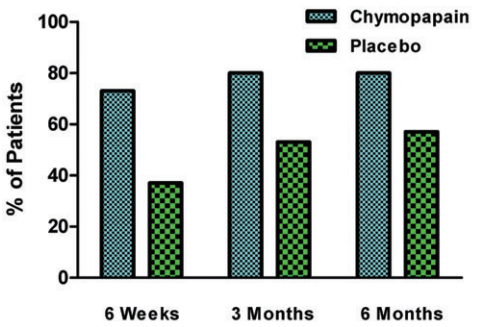


Fig. 1. Perception of success by patients in a placebo-controlled study of intra-disc injections of chymopapain for sciatica.[10]

between complementary and alternative medicine and conventional medical treatment enables the most holistic and effective management of the patient' and 'Hospital centres can easily adopt such a program by employing a health educator or wellness coach to provide supportive integrative services'.

A similar circumstance applies to a study by Eremin *et al*.[9] that looked at relaxation training and guided imagery for patients undergoing multi modal therapy for breast cancer. The control group received 'standard information and support', whilst the intervention group the patients 'were taught progressive muscular relation and cue-controlled relaxation training' as well as receiving advice about guided-imagery, which included 'visualising host defences destroying tumour cells'. The point is that the standard therapy is not always an appropriate control group. Allowances have to be made for placebo effects. Incidentally although it was labelled a 'clinical trial' the end points were laboratory measures of immunity. Claims of benefit through 'improved immunity' are the stuff of late night tele-commercials rather than surgical journals.

The use of an invasive placebo, such as the intra-arterial perfusion of saline, may raise ethical issues. Nevertheless, if the ethical requirements can be satisfied, such studies may provide interesting results. Figure 1 demonstrates that patients' perceptions of success can be maintained well after they have received a placebo. The unravelling of such effects over time is intriguing. Interventions that have a placebo response include: pulsed magnetic field therapy for tibial non-union, 'laser' therapy for soft-tissue injuries, extracorporeal lithotripsy for gallstones, and consultations with enthusiastic surgeons.[11]

The allocation of patients into groups

The randomization of patients into groups only achieves one thing with certainty – it ensures that there has been an unbiased allocation of patients. It eliminates any influence, either subconscious or deliberate, that the investigators might have on the allocation of the interventions.[12]

Randomization can result in groups with imbalances in a major risk factor, which is referred to as random confounding or pre-therapeutic bias. The ability of randomization to deliver an equitable dispersion of both known and unknown risk factors depends upon the sample size. The ability of chance to result in an uneven dispersion of risk factors diminishes as the sample size increases. However, even apparently large numbers of patients may be inadequate. A celebrated example was the University Group Diabetes Program (UDPG) study that dismissed a clinically and statistically significantly difference in risk factors with the comment that 'all in all, the luck of the draw was not too bad'.[13]

> The aim of randomization is to prevent the biased allocation of patients.
> Randomization does not guarantee that the groups have the same degree of risk.
> Be suspicious of clinical trails that have the same number of patients in each group in the absence of a blocked randomization technique

Valid randomization processes include the use of tables based on random numbers, computer generated 'pseudo-random' numbers, or the toss of coin. Systematic allocation methods – such as alternate allocations, allocations based on admission or birth dates, and allocations based on medical record numbers – violate the requirement that all patients have the same chance of receiving the intervention. They are not genuine forms of randomization.

The randomization schedule is a list of the interventions to be used to assigned patients into groups. It is also called the 'assignment schedule', 'allocation sequence', or 'randomization list'. The simplest randomization schedules consists of a simple list. For example, in a trial where the interventions are 'A' and 'B' the randomization schedule might be: B, A, A, B, A etc. When 20 patients are randomised, the chance of an 8:12 split, or worse, is about 50%, but if 100 patients are studied the probability is only 5%. Even with very large numbers of patients, it is unusual for a genuine simple randomization technique to result in an equal number of patients in each group. A disproportionate number of randomized studies report equal numbers in the group.

The simplest form of blocked randomization occurs when the allocation schedule contains subgroups that each contain an even distribution of the allocations. It might contain two lists, usually referred to as 'cells' each containing 10 allocations (5 for each group):

A, A, B, A, B, A, B, A, B, B
B, A, A, B, A, B, A, B, B, A

Blocked randomization ensures that the groups have similar sample sizes. Note that the last allocation in each cell is not at random, but this is an acceptable lapse of purity. Blocked randomization schedules take more complex forms when used for proportional assignments e.g., 2:1 rather than 1:1 allocations (permuted block design).

Stratified randomization is a technique for allocating patients into subgroups based on known risk factors. It is intended to result in a good balance between the groups of factors that might influence the main outcome events. In this process, each risk factor is divided into strata that contain their own randomization schedule. For example, the use of two gender strata (male, female), three age strata

© 2010 The Author
Journal compilation © 2010 Royal Australasian College of Surgeons

(20–49 years, 50–69 years, and 70 plus years), and four choices for a severity of disease classification would result in a total of 24 ($2 \times 3 \times 4$) cells. Use of a blocked randomization schedule for each of the cells helps to guard against gross imbalances. However, stratified randomization will fail to achieve its objective if the number of patients in each cell is small. This happens when a small sample size is combined with an excessive number of strata. As a general rule, stratification may improve power for trials with less than 400 patients, but only when the stratification factors have a large effect on outcome. The gains from stratified randomization are often small and the complexity of the process increases the risk of errors.

Baseline characteristics

The baseline is the initial time point in a clinical study before patients are exposed to the intervention. Readers should be interested in the characteristics of the patients at baseline for two reasons. First, you need to be convinced that the patients mirror your clinical practice ('generalizability'). Second, comparative studies need to convince you that each group had the same chance of a successful outcome at the start of the study (baseline equivalence).[14]

> At the end of a study it is assumed that all of the differences between the groups are due to the intervention

The classic 'Table 1' in comparative studies is a table documenting the baseline characteristics of the patients in each of the groups. It should show that the intervention and control groups are similar in terms of demographic variables (age and gender) and key prognostic variables (risk factors). When the intervention is an operation, it may include information about pathology or operative findings.

The risk factors must relate to the study endpoints. The aim is to prevent 'confounding' – a situation where the effect of an intervention is distorted by another factor. There are few situations where there are more than a handful of useful risk factors. Some authors try to build-up their article by providing spuriously long lists of baseline factors. It is not unusual to see lists that contain over 20 items. Most endpoints only have a few independent risk factors, and there is a limit to the amount of useful demographic data.

It is inappropriate to use P values to measure the extent of baseline equivalence.[15] One reason for this is that failure to detect a meaningful difference may be due to an inadequate sample size with a high probability of a type II (false-negative) error. Another reason is that P values are not indicative of the magnitude of difference between groups, they only concern the reliability of the observed differences (these issues will be discussed in parts VIII and IX of this series).

Withdrawal from study

Patients may be withdrawn from study because of:

- Incorrect entry of patient into trial (the patient did not fulfil the entry criteria or had an exclusion criteria).
- Loss of patient motivation.
- Clinical reasons, such as concurrent illness or pregnancy.
- Loss to follow up.
- Death.

Subjects who 'drop out' of research studies are more likely to be poor compliers and are more likely to have experienced complications. Hence, analysis of prospective clinical studies should be performed on an 'intention to treat' basis, and include all of the patients according to their initial allocation, regardless of the incompleteness of their therapy. Herman *et al*.[16] surveyed 274 orthopaedic trials and found that only 35% claimed to use the intention-to-treat principle and one-half of these did not adhere to its strict requirements. Most of the violations were due to the exclusion of patients lost to follow-up. It is too harsh to assume that all patients lost to follow-up are failures. There is probably no good way of handling missing data, 'you don't know what you don't know', but it is reasonable to expect investigators to tackle the problem by trying several methods of adjustment in a 'what if' (sensitivity) analysis.[17]

Follow up

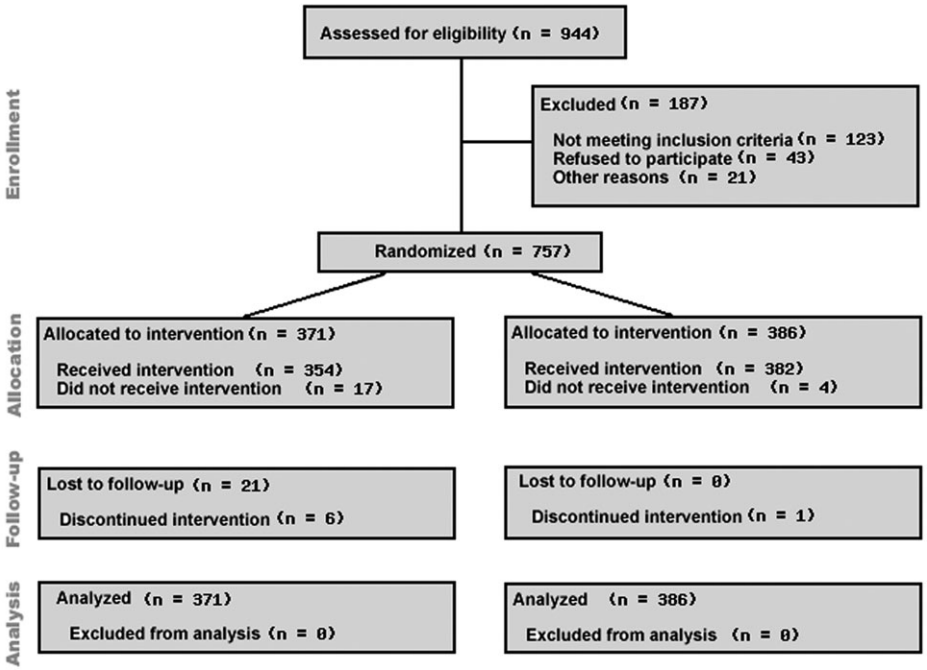
Some studies can be completed whilst the patients are still in hospital. However, most studies need to keep evaluating patients after they have left hospital. This introduces a number of problems, especially if the patients are not dependent on the clinic for continued care. It can be expensive and logistically difficult to continue to study patients after they have left hospital. Nevertheless, failure to do so can destroy the integrity of a study. For example, the rates of surgical site infections are often based on in-hospital data, when the median time of presentation of surgical site infections in contaminated surgery has been consistently reported to be in excess of 14 days.

The conclusions that are drawn at the end of a study should be consistent with the follow-up period. Frattareoli *et al*.[18] evaluated patients with 'early-stage prostate cancer'. One group received the 'usual care' and the other 'were encouraged to adopt a low-fat, plant-based diet, to exercise and practice stress management, and to attend group support sessions'. There were no significant differences between the prostate-specific antigen levels or velocities at the end of two years, but significantly more of the patients in the 'usual care' group had opted to start 'conventional prostate cancer treatment'. This led the authors to conclude that 'Patients with early-stage prostate cancer choosing active surveillance might be able to avoid or delay conventional treatment for at least 2 years by making changes in their diet and lifestyle'. There are many reasons for delaying conventional treatment – one of which might be the perception that they can avoid toxicity by assuming a 'healthy' lifestyle. Recommendations about interventions for prostate cancer should be based on long-term clinical outcomes. The question is not that the patients made a particular decision; but, rather, that it was the correct decision.

Although it is often desirable to follow patients for prolonged periods, which is usually difficult. But heroic efforts can result in success. In a prolonged clinical trial, Orloff *et al*.[19] followed-up

© 2010 The Author
Journal compilation © 2010 Royal Australasian College of Surgeons

Fig. 2. An example of a flow chart.



patients who presented with bleeding oesophageal varices due to cirrhosis 'for life'. The end result was a survival curve extending past 15 years that demonstrated the superiority of emergency porta-caval shunts over endoscopic sclerotherapy. This contrasts with the lack of information about the long-term outcome after bariatric surgery.

Investigators should report the number of patients who were lost to follow-up, and the reasons for this. The analysis of a study may be over-optimistic if some patients failed to complete the study. It is reassuring when investigators make clear statements about the nature and success of the review of patients. Unfortunately, some authors blur the issue. One technique is to mention in the methods section that all patients who failed to attend for review were approached in some manner, usually by mail or telephone. However, there is not followed by a declaration about the extent of such actions or their success. In addition, few studies with poor follow-up data outline the steps that were taken to facilitate the review of patients e.g., the inclusion of a commitment to allow follow-up in the consent form, the provision of details about close friends and relatives at the time of entry into the study, the ability to track patients who were lost to follow up.

The flow of patients

Whilst the classic 'Table 1' details the patients' characteristics at baseline, the classic 'Figure 1' is a flow chart of the patients. It has been recommended for use in clinical trials by the CONSORT Group (*www.consort-statement.org*) who provide authors with a digital template (Figure 2).

The principles are applicable to all clinical studies. Readers should not have to scratch around the text, or perform calculations, to try and find this information. Suspect the worst when authors do not declare exact numbers and only use percentages. They are either being stupid or deceitful; either way, continued attention is a waste of time.

When you flick through an article and see the classic 'Table 1' and 'Figure 1' then you are off to a great start. They, along with an illustration that summarises the main results, make it easy for readers to grasp the main features of a study. If authors fail to provide this trilogy of illustrations, and you think that the article may have merit, try to reconstruct the illustrations from the text. This is often a very informative exercise.

Key Points

- Patients pass through a number of selection pressures before they are included in a study.
- Entry, exclusion, and withdrawal criteria should be clear and unambiguous.
- Investigators tend to be over-optimistic about the numbers of patients that will enter a study.
- Results are always improved by omitting a control group.
- It is usual for the control group to receive the current 'gold-standard' therapy.
- The interpretation of clinical trials can be confounded by the 'placebo effect'.
- Look carefully at studies that include 'standard care' as the control group.
- Randomization ensures that there has been an unbiased allocation of patients.
- Randomization does not guarantee that the groups have the same degree of risk.
- Look for the classic Table 1 (baseline characteristics) and Figure 1 (flow of patients).

© 2010 The Author
Journal compilation © 2010 Royal Australasian College of Surgeons

- Subjects who 'drop out' of research studies are more likely to be poor compliers and to have experienced complications.
- At the end of a study it is assumed that all of the differences between the groups are due to the intervention.

References

1. Mishra VC, Brown J, Emberton M. Role of repeated prostatic massage in chronic prostatitis: a systematic review of the literature. *Urology* 2008; **71**: 731–5.
2. Kirkley A, Birmingham TB, Litchfield RB *et al*. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *N. Engl. J. Med.* 2008; **359**: 1097–107.
3. Lindström D, Sadr Azodi O, Wladis A *et al*. Effects of a perioperative smoking cessation intervention on postoperative complications: a randomized trial. *Ann. Surg.* 2008; **248**: 739–45.
4. Mitchell JR, Glover R, Kyle R. Bilateral internal mammary artery ligation for angina pectoris: preliminary clinical considerations. *Am. J. Cardiol.* 1958; **1**: 46–50.
5. Cobb LA, Thomas GI, Dillard DH, Merendino KA, Bruce RA. An evaluation of internal-mammary-artery ligation by a double-blind technique. *N. Engl. J. Med.* 1959; **260**: 1115–8.
6. Hall JC, Platell C, Hall JL. Surgery on trial: an account of clinical trials evaluating operations. *Surgery* 1998; **124**: 22–7.
7. Centa ID, Coggia M, Cochennee F, Alfonsi P, Javerlait I, Goëau-Brissonnière O. Laparoscopic abdominal aortic aneurysm repair in octogenarians. *J. Vasc. Surg.* 2009; **49**: 1135–9.
8. Kshettry VR, Carole LF, Henly SJ, Sendelbach S, Kummer B. Complementary alternative medical therapies for heart surgery patients: feasibility, safety, and impact. *Ann. Thorac. Surg.* 2006; **81**: 201–6.
9. Eremin O, Walker MB, Simpson E *et al*. Immuno-modulatory effects of relaxation training and guided imagery in women with locally advanced breast cancer undergoing multimodality therapy: a randomised controlled trial. *Breast* 2009; **18**: 17–25.
10. Fraser RD. Chymopapain for the treatment of intervertebral disc herniation: a preliminary report of a double blind study. *Spine* 1982; **7**: 608–613.
11. Johnson AG. Surgery as a placebo. *Lancet* 1994; **344**: 1140–2.
12. Hall JC, Hall JL. Randomisation in surgical trials. *Surgery* 2002; **132**: 513–8.
13. Feinstein AR. Clinical Biostatistics – XXIV. The role of randomization in sampling, testing, allocation, and credulous idolatry. *Clin. Pharmacol. Ther.* 1973; **14**: 1035–51.
14. Hall JC, Hall JL. Baseline comparisons in surgical trials. *ANZ J. Surg.* 2002; **72**: 567–9.
15. Senn S. Testing for baseline balance in clinical trials. *Stat. Med.* 1994; **13**: 1715–26.
16. Herman A, Botser IB, Tenenbaum S, Chechick A. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. *J. Bone Joint Surg. Am.* 2009; **91**: 2137–43.
17. Salim A, Mackinnon A, Griffiths K. Sensitivity analysis of intention-to-treat estimates when withdrawals are related to unobserved compliance status. *Stat. Med.* 2008; **27**: 1164–79.
18. Frattareoli J, Weidner G, Dnistrian AM *et al*. Clinical events in Prostate Cancer Lifestyle Trial: results from two-years of follow-up. *Urology* 2008; **72**: 1319–23.
19. Orloff MJ, Isenberg JI, Wheeler HO *et al*. Systemic encephalopathy in a randomized controlled trial of endoscopic sclerotherapy versus emergency portacaval shunt treatment of acutely bleeding esophageal varices in cirrhosis. *Ann Surg* 2009; **250**: 598–610.

John C. Hall, MS, DS, FRACS
Surgery (Royal Perth Hospital),
The University of Western Australia,
Perth, Western Australia, Australia
john.hall@uwa.edu.au

doi: 10.1111/j.1445-2197.2010.05355.x

© 2010 The Author
Journal compilation © 2010 Royal Australasian College of Surgeons