Evaluating surgery

Jonathan L Meakins

J R Soc Med 2007;100:357–359

In this issue of the *Journal* (*JRSM* 2007;100:387–390), Professor McPherson recounts and comments upon the remarkable book *Costs, Risks and Benefits of Surgery*, published 30 years ago by Bunker, Barnes and Mosteller.¹

McPherson was a participant in the writing process which was, for academic circles, remarkably egalitarian. The hierarchy of rank did not get in the way of the generation of ideas or their eventual articulation in print. He has noted its listing by Black and Neuhauser as one of the 26 books to have changed health services and health policy.² Of the 135 books which were evaluated, only two were on surgical subjects: this and Codman's report of his private hospital's end results.³

Codman paid a penalty for his insistence on publishing surgical outcomes; it was very unpopular with Boston surgeons and he lost his post at the Massachusetts General Hospital. The link between the two books is clear, in that both insist on knowing outcomes; although *Costs, Risks and Benefits* extends those concepts into the evaluation of and innovation in surgery, as well as quality of life as an important outcome. The books therefore look at two types of outcome. Codman's are largely surgeon-centred, focused on complications and mortality rates, while Bunker *et al.* devoted a chapter to quality of life outcomes. In the section they wrote on established procedures, the principle of patient-centred outcomes surfaces repeatedly.

When referring to surgeon-centred and patient-centred outcomes, perhaps a hernia repair is the most useful operation to examine. Surgeons are primarily concerned with recurrence rates (generally the focus of all studies, randomized control trials (RCT) and case series), which have a propensity to be low but to increase over time. Each surgeon tends to experience few recurrences of hernia, both because time passes before such an event occurs and because a patient will often visit a second surgeon rather than the initial operator. From the patient's point of view, however, recurrences are often of less importance than adverse effects such as chronic pain, testicular atrophy, wound complications or paresthesias.⁴ Exchanging a minimally symptomatic or an asymptomatic hernia for a chronic pain or paresthesia seems like a poor deal for the patient. More recent evidence

now indicates clearly that there are significant patient costs to having a hernia repaired. About 20% of patients have some long-term complaint in association with their operation.^{4,5} Bunker *et al.* questioned the value of prophylactic surgery in general and for hernia in particular—an issue which has only recently been resolved.^{5,6}

One of the questions asked by Bunker *et al.* was, 'Why were there such variations in clinical practice and procedures rates within and between countries?' One would have assumed that with adequate evidence, regional variation would diminish. A point that comes through in the book is that there was indeed a lack of evidence. A key objective was to highlight the importance of evaluating procedures and innovation and understand short- and long-term outcomes.

In the section on innovation and evaluation, Bunker *et al.* look largely at failed procedures; as one reads about surgery for ptosis or constipation, or of the endocrine glands, one wonders how these procedures could possibly have been thought useful. These examples demonstrate the difficulties arising when strong characters forcibly and articulately present a point of view. It is the power of the expert, or 'eminence-based practice'. Yet it was evidence-based medicine and RCTs which eventually stopped the practice of gastric freezing for duodenal ulcer or the use of the internal mammary artery ligation as management for angina pectoris.

Are we doing better 30 years on? There have been vociferous complaints about the quality of surgical research.⁷ The presumption has always been that RCTs are a solution to all clinical therapeutic questions. Within the domain of surgery, and indeed the broader domain of procedure-based medicine, this has been challenged.^{8–10} Examples of evolving clinical benefit without RCTs are found in the management of burns, drainage of abscesses, management of trauma and bleeding, transplantation and resection as the only possible cure for many cancers. The drainage of subphrenic abscesses tells an interesting story. It was difficult to diagnose a subphrenic abscess before the advent of computed tomography, but they are clinically very important. The aphorisms 'pus somewhere, pus nowhere, pus under the diaphragm' and 'never let the sun set on an undrained abscess' highlighted the importance of the management of this problem. A number of drainage

procedures were defined, including minimally-invasive approaches using sophisticated tissue planes. In 1981, Gerzof *et al.* demonstrated that these abscesses could be drained percutaneously using interventional radiological techniques.¹¹ The surgical community was up in arms at this change in practice, yet clearly the better minimal access approach was effective and good for patients. An RCT not only was not done but was not necessary. The professionals relaxed and percutaneous drainage of collections became standard practice. A new technique with a large improvement in outcomes is much more easily accepted than one which provides only a small improvement.

Outcomes after cardiac surgery, major vascular operations and oncological procedures, amongst others, have continuously improved without trials, secondary to improvements in patient assessment, surgical technique, anaesthetic approaches, and global peri-operative care—despite operating, generally speaking, on an older and sicker population of patients. The improvement often results from the integration of evidence-based therapies demonstrated in other settings into the treatment pathway. Perioperative and operative management are constantly evolving with interactive changes measured against historic performance.

The description of costs, risks and benefits in the 1970s was almost exclusively from a single ‘procedure’ point of view—surgery. Much has changed in clinical medicine in the last 30 years. Surgery continues with the same issues around evaluation and innovation. However, the tremendous advances in technology applied to medical practice have introduced ‘procedures’ into almost every specialty. Fibre-optics have allowed endoscopes to be introduced into virtually every orifice of the body; many open operations have been converted to closed procedures; and most medical specialties now have a repertoire of procedures once considered operations. The revolution in imaging has created a new specialty—Interventional Radiology—with a host of percutaneous therapeutic and diagnostic opportunities in body cavities and via intravascular routes. The abundance of options has led to difficulties in initial evaluation, intradisciplinary arguments about value, and intense turf battles between specialities and disciplines.

The classic RCT can resolve some of these problems; however, a selection of evaluative tools and study designs are now required to address the dramatic pace of technological change and evolving clinical applications. Operations—procedures would be a more comprehensive term—are no longer performed exclusively by surgeons. The evaluation toolbox needs to cover all procedure-based medical practices, which now encompasses most specialities.

Thirty years on, many of the problems outlined in *Costs, Risks and Benefits* remain, but they are no longer exclusively

surgical. Bunker *et al.* made four recommendations: two will be addressed here, and are the specific subjects of meetings which have already been initiated.

Recommendation four from *Costs, Risks and Benefits* was, ‘Information on outcomes as well as costs of medical care should be routinely formulated in a manner suitable for presentation to the public’. This reflects on a patient-centred world and links to patient participation in clinical decision making. In light of the universal shortening of surgical training and education time as a result of the limitation of work hours in both Europe and North America, the duration of training and education will shorten. As patients become better educated and demand a role in their own care; decision making will have to be taught as well as learned by osmosis. A series of workshops on decision making sponsored by the *British Journal of Surgery* and organized by the Nuffield Department of Surgery and the Royal College of Surgeons was initiated during February 2007. Two more are in the planning stages.

Recommendation one from *Costs, Risks and Benefits* was, ‘Appropriate studies of the effectiveness of surgical treatment should be carried out for selective conditions, particularly those where uncertainty leads to professional disagreement’. This recommendation has more importance in light of expanding procedure-based practice. Increasingly, technology is allowing different disciplines to manage the same clinical condition: coronary artery disease ischaemia may be dealt with by cardiology or cardiac surgery; cerebrovascular aneurysms by interventional neuroradiology or neurosurgery; liver metastases by surgery or percutaneous ablation by a variety of techniques; common duct stones by surgery or endoscopy. What evaluative techniques warrant the new approach being tested against the traditional? Independent of turf battles, without the evidence how can proper clinical decisions be made? What evidence will be adequate? Is the RCT the only solution? Clearly it is not, although it is just as clearly the gold standard for comparison of two therapeutic approaches.

The arrival of laparoscopic cholecystectomy has led to a dramatic proliferation of new procedures and new approaches to open surgery. Progress has been astonishing, and many procedures using the laparoscope or other minimally invasive techniques have been established without formal RCTs. Indeed, calls for an RCT have in some instances delayed the adoption of techniques that seem, on the basis of observational studies, to be clearly in patients’ best interests. A good example is laparoscopic donor nephrectomy, which was shown in observational studies to provide kidneys which were as good as those provided through open nephrectomy, with exactly the same or better complication rates and a high level of patient satisfaction.^{12,13}

Indeed, the technique substantially increased the donor pool, as individuals were willing to provide a kidney in the absence of the large incision and its painful consequences.

A recent comment in the *Lancet* entitled 'Minimising risk in first-in-man trials' looks only at issues in association with pharmacology;¹⁴ however, first-in-man trials are happening, if not daily, at least on a regular basis in the operating, endoscopy and radiology suites of the western world, and the toolbox for their evaluation and subsequent adoption needs to be defined. Given this situation, a series of colloquia have been established in Oxford at Balliol College to clarify the various methods for evaluating surgical procedures. Collectively, we recognized that RCTs are the gold standard but that other techniques are required where the RCT is not feasible. The bland repeated statement that all therapies are suitable to an RCT is not useful. A definition of when other techniques are suitable and a description of the toolbox applicable to these techniques is important and urgently required. The long-term value of these approaches will be to determine, in a disciplined and defined manner, and using a variety of techniques, the ways in which we can assess the rapidly evolving procedure-based approaches to the management of clinical problems. Operations by surgeons and procedures done by other disciplines entail huge resources within the NHS and, indeed, the health care systems of the western world. Their correct evaluation and suitable techniques to define those of value are of considerable importance.

In March 2007, the Balliol colloquia reviewed the current state of play in evaluation of surgical procedures, complicating issues in surgery, the differences between studying drugs and procedures, and problems associated with and solutions for the surgical RCT. Extensive review of non-randomized intervention studies and discussion of designs applicable to procedural intervention occupied a full day. The next colloquium will be in September 2007 and, based upon the record of the March meeting, will study designs suitable for procedural evaluation and define the experimental toolbox. The attendees are a variety of clinicians, methodologists, statisticians and epidemiologists, all embracing different points of view. The meeting structure is very similar to that described in the seminar series in *Costs, Risks and Benefits*; however, because the participants come from all corners of the UK as well as

continental Europe and North America, having meetings every other week is not possible. Three meetings of three days each are anticipated, with very short presentations and wide-ranging discussions by all participants. All comments are recorded and the first meeting integrated into an overall plan defining the content of the second meeting, which will in turn define the content of the third meeting. The results of these exercises will be published. It is expected that the evaluative instruments—from case reports to RCTs—will be defined and their indications clarified, facilitating evaluation of surgery specifically and procedural therapies in general. Our patients and the health care system can only benefit.

REFERENCES

- 1 Bunker JP, Barnes BA, Mosteller F. *Costs, Risks and Benefits of Surgery*. New York: Oxford University Press, 1977
- 2 Blach N, Neuhauser D. Books that have changed health services and health care policy. *J Health Serv Res Policy* 2006;**11**:180–183d
- 3 Codman EA. *A Study in Hospital Efficiency as demonstrated by Case Reports of the First Five Years of a Private Hospital*. Boston, 1916
- 4 Hawn MT, Itani KM, Giobbie-Hurder A, McCarthy M, Jonasson O, Neumayer LA. Patient-reported outcomes after inguinal herniorrhaphy. *Surgery* 2006;**140**:198–205
- 5 Neumayer L. Is the presence of an inguinal hernia enough to justify repair? *Ann Surg* 2006;**244**:174–5
- 6 Fitzgibbons RJ, Giobbie-Hurder A, Gibbs JO, *et al.* Watchful waiting versus repair of inguinal hernia in minimally symptomatic men: a randomized clinical trial. *JAMA* 2006;**295**:285–92
- 7 Horton R. Surgical research or comic opera: questions but few answers. *Lancet* 1996;**347**:984–5
- 8 Meakins JL. Innovation in Surgery: the rules of evidence. *Am J Surg* 2002;**183**:399–405
- 9 McCulloch P, Taylor I, Sasaka M, *et al.* Randomised trials in surgery: problems and possible solution. *BMJ* 2002;**324**:1448–51
- 10 Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007;**334**:349–51
- 11 Gerzof SG, Robbins AH, Johnson WC, Birkett DH, Nasbeth DC. Percutaneous catheter drainage of abdominal abscesses: a five-year experience. *N Engl J Med* 1981;**305**:653–7
- 12 Tooher RL, Rao MM, Scott DF, *et al.* A systematic review of laparoscopic live-donor nephrectomy. *Transplantation* 2004;**78**:404–14
- 13 Tooher R, Boulton M, Maddern GJ, Rao MM. Final report from the ASERNIP-S audit of laparoscopic live-donor nephrectomy. *ANZ J Surg* 2004;**74**:961–3
- 14 Hemelaar J. Minimising the risk in first-in-man trials. *Lancet* 2007;**369**:1496–7