Published in final edited form as:

Lancet. 2009 September 26; 374(9695): 1097–1104. doi:10.1016/S0140-6736(09)61086-2.

Challenges in evaluating surgical innovation

Patrick L Ergina, MD,

Department of Surgery, McGill University Health Centre, Montreal, QC, Canada; Oxford International Programme in Evidence-Based Health Care, Balliol College, Oxford University, Oxford, UK

Jonathan A Cook, PhD,

Health Services Research Unit, University of Aberdeen, Aberdeen, UK

Prof Jane M Blazeby, MD,

Department of Social Medicine, University of Bristol, Bristol, UK

Isabelle Boutron, MD,

Centre for Statistics in Medicine, University of Oxford, Oxford, UK

Prof Pierre-Alain Clavien, MD,

Department of Surgery, University Hospital Zurich

Prof Barnaby C Reeves, DPhil, and

Department of Clinical Science

Christoph M Seiler, MD

Department of General, Visceral and Transplantation Surgery, University Hospital of Heidelberg, Heidelberg, Germany

for the Balliol Collaboration*

Abstract

Research on surgical interventions is associated with several methodological and practical challenges of which few, if any, apply only to surgery. However, surgical evaluation is especially demanding because many of these challenges coincide. In this report, the second of three on surgical innovation and evaluation, we discuss obstacles related to the study design of randomised controlled trials and non-randomised studies assessing surgical interventions. We also describe the issues related to the nature of surgical procedures—for example, their complexity, surgeon-related factors, and the range of outcomes. Although difficult, surgical evaluation is achievable and necessary. Solutions tailored to surgical research and a framework for generating evidence on which to base surgical practice are essential.

Correspondence to: Dr Patrick L Ergina, McGill University Health Centre, Royal Victoria Hospital, Division of Cardiothoracic Surgery, 687 Pine Avenue West, S8.76B, Montreal, QC H3A 1A1, Canada patrick.ergina@mhcc.mcgill.ca.

*For members see *Lancet* 2009; **374**: 1089–96

Contributors The writing group of this paper (PLE, JAC, JMB, IB, P-AC, BCR, CMS, and the Balliol Collaboration) agreed to the form and content at a meeting on April 3, 2009. PLE and JAC wrote the initial outline, the drafts, and coordinated and edited contributions from the other authors. Members of the Balliol Collaboration reviewed the circulated drafts and made comments and contributions.

This is the second in a **Series** of three papers on surgical innovation and evaluation

Conflicts of interest We declare that we have no conflicts of interest.

Introduction

Evaluation of a therapeutic, procedure-based intervention presents several methodological and practical challenges for the surgical research community. Few, if any, of these challenges apply only to surgical procedures; many arise during the assessment of other non-pharmacological interventions, such as interventional radiology, technical procedures and devices, rehabilitation, behavioural interventions, and psychotherapy.¹ However, what is arguably unique to surgery is the way in which many of these challenges coincide. Perhaps this situation leads many surgeons to view randomised controlled trials (RCTs)—although theoretically advantageous—to be too difficult and impractical to undertake, and at worst, irrelevant to their practice because of concerns about generalisability.² Most of the same challenges also affect non-randomised studies and, in some cases, to a greater extent. Despite the barriers, an RCT remains the best possible study design for the assessment of therapeutic interventions.

This report, the second of three papers on surgical innovation and evaluation, presents the conclusions of a meeting held by the Balliol Collaboration on April 3, 2009. By identifying many issues related to surgical research and deconstructing them into constituent methodological parts, we targeted several important areas to develop guidance for appropriate, evidence-based surgical practice. Here, we discuss the challenges related to study design of surgical research and the challenges related to the nature of surgical interventions. Recommendations for improvement and solutions are presented in the third report in this Series.³

Challenges related to study design

Randomised controlled trials

RCTs are considered the gold standard for establishing safety and efficacy of an intervention. Despite calls for surgical research to be more rigorous, the overall frequency of RCTs has been consistently low since the 1970s.⁴ Large, high-quality RCTs have been done in a variety of surgical specialties, but those of the surgical procedure itself are less common. Most surgical RCTs have focused on other aspects of the intervention, such as anaesthesia or pharmacological interventions, in preoperative and postoperative care.⁵ There are several types of RCT that can be used for different aims.⁶ A useful distinction can be made between explanatory trials, which seek to assess whether an intervention can work, and pragmatic trials, which seek to inform clinical decision making. Pragmatic trials are needed to ensure surgical practice is based on evidence. Some criticisms of surgical trials are misplaced and reflect misunderstanding of a trial's aim. Specific challenges to the planning and conduct of randomised trials comparing a surgical intervention with different types of comparator are summarised in the table. Many of the issues raised might need different solutions, depending on the aim of the study.

A particularly difficult question in the assessment of a new surgical intervention is whether an RCT is necessary and, if so, when the first one should be done. Conceptually, there are few arguments against doing RCTs early in development, although a further assessment might be appropriate. For a new surgical intervention, it can be difficult to decide when to shift from an early exploratory stage of development to a formal investigation. If done too early, the constraints of an RCT could obstruct innovation, and if too late, equipoise could be lost. Another consequence of an early RCT is that the definitive technique might not be fully refined; the subsequent study outcome then reflects the stage of development and learning, and not the therapeutic effect of the intervention.¹¹ Additionally, restricting a new procedure to an RCT might be impractical in the absence of regulation that prevents surgeons offering the intervention to patients outside the trial.¹²

When two interventions have different benefit-to-harm profiles, patients and surgeons might strongly prefer one intervention. Strong preferences might lead patients and surgeons to decline trial participation, making trial recruitment more difficult if not impossible. This situation is often intractable when the preferred intervention is widely available. The strength of a patient's preference partly depends on the comparator used in the trial (table). Trial designs that seek to evaluate preferences have been proposed.^{13, 14} Possible comparisons are a new procedure versus a sham (placebo) procedure; similar but distinguishable procedures; substantially different procedures; or surgery versus non-surgical treatment, such as a medical treatment, participative intervention (eg, rehabilitation), or watchful waiting. Research to assess surgical innovations in emergency and paediatric settings are perhaps particularly susceptible to preferences that preclude randomisation. Breast surgery is the archetypal example (difficulty of randomising patients to segmental mastectomy or mastectomy).¹⁵

Response to uncertainty has also been suggested as an explanation why surgeons might be unwilling to take part in surgical trials. Compared with physicians, surgeons might be less tolerant of uncertainty about the effectiveness of alternative treatments, affecting their participation in RCTs and thereby making surgical trials more difficult to undertake.¹⁶ Previous negative experiences and perceived threat of litigation might make some surgeons reluctant to submit parts of their practice to evaluation. A feasibility study of patients' willingness to participate in surgical and oncology trials found a low level of willingness because of a stated dislike for randomisation, and a desire to make their own decisions about the selection of the intervention.¹⁷ Preferences of both surgeons and patients might rarely be based on existing evidence. Patients need to be provided with sufficient information to make informed decisions, which has not always been the case.¹⁸ Qualitative research can provide insights into the recruitment process and might enable greater participation in RCTs.¹⁹

Randomisation should be done as close to the time of the intervention as possible to reduce the possibility that the allocated intervention will not be delivered because of strong preferences, knowledge of allocation assignment, cancellations, or clinical events before the procedure.²⁰ However, randomisation needs to be sufficiently early for the patient and surgical team to be adequately prepared. In the case of two surgical procedures, randomisation can often be done in the operating theatre, for example, by use of a telephone or web-based randomisation service.²¹ The challenge is exaggerated when substantially different (eg, surgical vs pharmacological) interventions are compared and participants have to be told their allocation in advance of receiving the intervention, or when the new procedure is available outside the trial. Irrespective of the timing of randomisation, a surgeon might decide that a surgical procedure is inappropriate, impossible, or unsafe after randomisation. The Spine Patient Outcomes Research Trial (SPORT),¹⁰ which compared surgery and non-surgical management in low back pain, reported substantial crossing over of patients in both directions. Although the principle of intention to treat provides the preferred analysis framework for dealing with crossing over, application of the results of such an analysis to other settings can be difficult (for example, if crossovers are predominantly from the new treatment to the conventional treatment).

Absence of masking can lead to several forms of bias: performance bias (surgeons, other caregivers, or patients choosing concurrent interventions depending on their allocation); attrition bias (differential withdrawal from follow-up); and detection bias (differential outcome assessment).²² Masking of surgeons, patients, and other caregivers is difficult and often impossible in surgical trials; nevertheless, innovative methods of masking are available.²³ In a comparison of laparoscopic and small-incision cholecystectomy, bloody bandages were used to blind patients and other caregivers.²¹ Sham, or placebo surgery, in which the surgeon mimics the intervention, has been used to assess arthroscopic surgery for

osteoarthritis of the knee and stem-cell therapy for Parkinson's disease.^{7,24} The use of placebo surgery is controversial, and has been restricted to cases where a suitable comparator was not available or the placebo surgery had limited risk.²⁵ Although masking of the surgeon and patient is difficult, it should be possible to blind the clinical assessment of outcomes (though seldom done).²⁶ If patients cannot be masked to treatment assignment, some outcomes can be susceptible to bias, especially patient-reported outcomes.

The principle of random allocation of participants to surgeons with expertise of different procedures—ie, an expertise-based design (an intrinsic feature of a comparison between a surgical procedure and a non-surgical treatment)—has been proposed for the comparison of two surgical procedures.²⁷ Similar to cluster randomisation, this design protects against contamination and allows surgeons with strong preferences to take part. However, this design brings its own challenges: more surgeons are required, the comparison could be confounded by the characteristics of surgeons who prefer one technique, and the logistics of shared waiting lists across surgeons are formidable.

A tracker trial design has been proposed to reflect and incorporate the difficulty of incremental and stepwise innovation during assessment of a surgical procedure.¹² In this design, modifications of the surgical technique during the progress of the trial are allowed, recorded, and subsequently tracked in the statistical analysis. Variations in the randomisation scheme, such as adding a new treatment, are also allowed. In principle, the full development of an innovation can be assessed in a single study. Although conceptually attractive, tracker trials would be very challenging in practice.

Non-randomised studies

As previously mentioned, several factors contribute to make RCTs of surgical procedures difficult and, in a few cases, impossible. For example, lesser surgical innovations might have such a small effect on a serious but rare outcome (eg, mortality) as to make an RCT evaluation prohibitively large to achieve adequate statistical power. Historically, most advances in surgical knowledge have been accepted on the basis of non-randomised studies.²⁸ Surgical interventions such as heart, liver, kidney, and lung transplantation are established therapies in developed countries.²⁹ None of these procedures has been validated with RCTs, and it is generally regarded unethical to do so in view of the apparent benefit.³⁰ Other advances have been identified through observational studies, or even anecdotes, because of dramatic effects, where biases are unlikely to be so severe that they could account for the findings.³¹

Early exploratory cases of new procedures are likely to be reported as case reports or case series.³² Large cohort observational studies have been critically and extensively used to develop and validate risk assessment for surgical therapies, to monitor safety in practice, to identify treatment effects (adverse or beneficial) that might not have been looked for or detected in original studies, and to estimate treatment effects when RCTs were deemed impracticable (eg, rare events, observations far in the future).³³ When RCTs are not feasible, it is essential to undertake high-quality non-randomised studies.³⁴ A dichotomy between randomised and non-randomised studies is somewhat artificial since both designs can provide different and complementary evidence.³⁵ For example, a non-randomised investigation of long-term and rare safety outcomes could be done alongside an RCT.

Overall, most surgical studies are non-randomised and often retrospective; their quality is also very variable and often poor.⁴ Prospective comparative designs are substantially more useful than case series, which are over-represented in surgical publications. An important driving factor behind non-randomised studies is that they are easier to undertake than RCTs, and increasingly so with electronic data collection and standardised databases.³⁶ However, a

lack of appropriate planning and poor data quality (missing data for important risk factors, inconsistencies, and the absence of key diagnostic and operative details) are common setbacks that tend to undermine the validity of results from non-randomised studies.

Protocol-driven studies, which account for all cases and have accurate and informative clinical data, are needed. More effort should be focused on data collection to reduce bias caused by incomplete data or unmasked outcome assessment, as is the case in RCTs. However, even well-designed non-randomised studies face many of the difficulties associated with RCTs—for example, the existence of a learning curve. Accounting for any pretreatment differences between intervention groups is also a particular concern in non-randomised studies.³⁷ Rigorous prospective design and data collection provide some protection against biases.³⁸ The use of any statistical adjustment (such as propensity scores) to overcome potential confounding effects will only have merit if informed by comprehensive clinical understanding of the condition and its risk factors. The detailed patient data needed to undertake such an analysis is rarely available.

Finally, causal inferences established in a non-randomised study are judged weaker than those identified in an RCT and need cautious interpretation.³⁹ There are several examples of established surgical practices—previously validated with non-randomised studies—that have been discontinued after testing in a large RCT (eg, extracranial to intracranial bypass,⁴⁰ carotid endarterectomy,⁴¹ lung volume reduction surgery⁴²). Since current advances have been more subtle, the need for RCTs should increase—ie, the smaller the difference in outcome, the greater the need for an RCT.

Challenges related to the nature of surgical interventions

Complexity of surgical procedures

We need to recognise that many surgical interventions are complex and require appropriate evaluation.⁴³ Surgical interventions, like other non-pharmacological interventions such as therapist-based and educational interventions, consist of several components that cannot be separated.⁴⁴ This situation contrasts with most pharmacological interventions, which can be readily defined and standardised. Although the surgical procedure itself requires attention, a surgical intervention can depend on many health-care professionals and involve other aspects of health-care delivery in ways that a pharmacological intervention does not (figure).

A surgical procedure is mainly delivered by a surgeon and is affected by characteristics such as surgical skill, decision making, preferences, and experience. The delivery of a surgical intervention also depends on the other members of the team (eg, anaesthetists, nurses, technicians) and preoperative and postoperative management (eg, emergency department, imaging services, postoperative recovery ward, intensive care, and rehabilitation programmes). This complexity often receives little recognition in the design of surgical studies. Indeed, its existence is sometimes used to criticise studies of surgical interventions for failing to control for potential confounding factors.⁴⁵

An example of a typical complex surgical intervention that consists of several interacting components is coronary artery bypass graft surgery (CABG). The aim of this procedure is to revascularise the myocardium by bypassing coronary arteries that are stenosed or blocked. Several steps constitute the surgical procedure: opening the chest; harvesting conduits; attaching (and later detaching) the heart–lung machine; undertaking the anastomoses; reanimating the heart; closing the chest. In the case of CABG, there is limited variation in technique between surgeons.^{46,47} However, there are many recognised variations in surgical strategy, such as off-pump CABG (avoidance of the heart–lung machine), minimally invasive approaches, and different choices of bypass conduits (eg, bilateral mammary

arteries, radial arteries). Some decisions are made intra-operatively (eg, whether additional grafts are needed) and will depend on the judgment of the individual surgeon. Other co-interventions might be used, such as antifibrinolytic agents, insulin, or hypothermia. Preoperative medical care (eg, coronary care unit/cardiology management, medical management of comorbidities, blood bank management), roles of other members of the surgical team (eg, nurses, anaesthetists, perfusionists), and postoperative care (eg, intensive care, acute and chronic cardiac rehabilitation) also vary and affect outcomes.⁴⁸ These supporting components vary between centres and are affected by infrastructure, staffing, and local policies.

Although an intervention needs to have a coherent aim (or function), different forms are often available.⁴⁹ The complexity, and potential variability, of a surgical intervention raises two difficult questions for the design of a surgical evaluation for which only general answers can be given. First, when is variation in form substantial enough to be worth assessing? Second, when investigating alternatives, how standardised should they be, in view of the complexity of the steps involved? Continuing the CABG example, does avoidance of the heart–lung machine warrant investigation? If so, how standardised should the off-pump CABG surgical strategy and other steps be? The effect on health services (eg, equipment resources, staff requirements such as training), the potential for a change in the balance of benefits and harms, or consensus among surgeons could justify assessment of alternatives. The degree of intervention definition and the level of standardisation of the new approach will depend on the stage of development and the aim of the evaluation. The amount of information that researchers need to record about the conduct of an intervention will depend on how an intervention is defined and the degree of standardisation sought. Very restrictive approaches could limit surgeon participation and might not be feasible in some centres.

Surgeon-related factors

As previously mentioned, attributes of the surgeon, such as surgical knowledge, previous training and experience, and inherent skills, will influence the delivery of a surgical intervention and lead to variability in practice and health outcomes. Variability can be expected irrespective of previous training and experience. Differences between surgeons interact with patients' differences, affecting the responses to operations. The expectation that all surgeons should attain the ideal, often high level of performance is unrealistic. Evaluations of surgical procedures should therefore be done in realistic settings.

The learning curve for a surgical intervention, whereby surgeons acquire expertise, poses an important challenge. Since the technical and functional success of a procedure is paramount, the early stages of assessment, and thus publication of results, tend to focus on complications.^{50,51} For example, the rate of bile duct injuries associated with laparoscopic cholecystectomy fell as the surgeons' experience increased.⁵² Proxies for operative expertise, such as duration of surgery and amount of blood loss, have been used to assess the impact of learning.⁵¹ The effect of the learning process on health outcomes is subject to debate and likely to vary between interventions. For complex operations (eg, radical prostatectomy and laparoscopic hernia repair), learning can continue over a very long time, perhaps hundreds of procedures.^{9,53}

Evaluation of a new surgical intervention versus an established control has been criticised, owing to a perceived imbalance of experience that favours the established comparator.⁵⁴ Some have sought to undertake studies that include surgeons who have completed their learning. This strategy is complicated by individual surgeons learning at different rates and the effect of external factors on the learning process itself.⁵⁵ Trial design could be modified to incorporate individual learning, and studies of surgical innovations should consider the effect of learning. Better recording of surgical training and the experience of participating

surgeons would be a step forward. Collection of comprehensive data on new interventions, which requires surgeons to document personal procedure-based learning, would allow a more informative assessment of surgical learning.⁵⁶

Surgical outcome evaluation

The key questions to address when planning a study of a surgical intervention are: what is the outcome, how should it be measured, who should assess it, and when? The quality assurance literature has used the terms structure, process, and outcomes as suggested aspects for measuring the quality of surgical care.⁵⁷ Traditionally, surgeons themselves have selected and assessed the outcomes, mainly focusing on short-term clinical measures of technical success and harm. However, such outcomes are often not standardised and therefore not reproducible, which hinders evaluation. For example, a systematic review showed that in 107 studies, there were 56 separate definitions of anastomotic leak at any site after gastrointestinal surgery, precluding comparison of leak rates between studies.⁵⁸ The absence of standardised (agreed upon) surgical terminology for the definition of clinical outcomes has long been recognised and this has led to the development of methods for grading and classifying deviations from the normal postoperative course, which have been tested, modified, and validated.⁵⁹ One proposed strategy is to use a validated therapy-oriented classification system for complications, which ranks adverse events by severity with avoidance of confusing terms (panel).^{59, 60} This system could be adjusted to match a clear and consensus definition of postoperative events within specific specialties of surgery.⁶¹

Although these surgeon-selected (or physician-centred) clinical outcomes (eg, mortality and morbidity rates) are very important to patients, evaluation of surgery needs to be widened to include the patient's perspective. Patients' perceptions—and thus reporting of symptoms and function—can differ from the surgeon's assessment; what patients judge as important (eg, social, emotional function) might be different from the issues of interest to surgeons. Therefore, studies of surgical interventions require assessment of both clinical and patient-reported outcomes. Typically, this information is captured in questionnaires assessing health-related quality of life. It can be difficult to decide which outcomes are best suited to a particular medical problem. Methods to select and incorporate assessment of health-related quality of life in trials are emerging and better performed studies will produce more reliable data. However, despite the recent interest in this area, there seems to be a gap between measuring health-related quality of life outcomes and using the information to change surgical practice.⁶² This division might occur because the surgical community does not understand the data or because clinical outcomes are considered paramount. Methods to accurately measure and interpret patient-related outcomes alongside clinical data are needed so that surgeons can effectively evaluate surgery and subsequently inform patients.

There are some situations in which patient-related outcomes are more important than clinical outcomes (eg, palliative surgery, functional outcomes after joint replacement surgery) and capturing these data within well-designed trials is essential. Developing core outcome sets with key clinical, technical, and patient-reported outcomes will help to facilitate the process. Methods to reach agreement about these outcomes have been developed in rheumatology.⁶³ Additionally, economic evaluation is crucial for efficient use of often limited resources.

Panel: Clavien–Dindo classification of surgical complications

Grade I

Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic, and radiological interventions.

Allowed therapeutic regimens are: drugs as antiemetics, antipyretics, analgesics, diuretics, electrolytes, and physiotherapy. This grade also includes wound infections opened at the bedside

Grade II

Requiring pharmacological treatment with drug other than such allowed for grade I complications. Blood transfusions and total parenteral nutrition are included

Grade III

Requiring surgical, endoscopic, or radiological intervention

- Grade IIIa: intervention not under general anaesthesia
- Grade IIIb: intervention under general anaesthesia

Grade IV

Life-threatening complication (including CNS complications)* requiring intermediate care or intensive-care unit management

- Grade IVa: single-organ dysfunction (including dialysis)
- Grade IVb: multi-organ dysfunction

Grade V

Death of a patient

Grading system proposed in 2004. The key concept of this scale was that objective severity of a complication could be defined by the treatment it provoked to reverse it, or death. *Brain haemorrhage, ischaemic stroke, subarachnoid bleeding, but excluding transient ischaemic attacks. From references ⁵⁹ and ⁶⁰ with permission.

A more comprehensive approach to studying surgical procedures is needed. This approach should use accurate, standardised clinical and patient-reported outcomes, recorded in real time, and whenever possible by an independent observer who is masked to treatment assignment. After the early development of surgical interventions, comprehensive assessment of outcomes is recommended for all other stages of development.³ This approach provides information to allow evidence-based comparisons between different interventions.

Additional challenges in surgery

Traditional master–student model

The traditional hierarchical system of surgery epitomises eminence-based medicine. This master–student apprenticeship tradition holds that the master has all the knowledge and skill and the student learns by observation and emulation. This approach can prevent new models and information from entering independent practice. Despite attempts to implement change with aggressive knowledge translation methods,⁶⁴ adoption of best practice guidelines in surgery remains poor without involvement of surgical opinion leaders.⁶⁵ This might help explain, in part, the slow acceptance of evidence-based surgery, and in particular, RCTs. Meakins' editorial introducing the first users' guide for evidence-based surgery did not appear until 2001,⁶⁶ well after the introduction of users' guides to the medical literature in 1993.⁶⁷

Lack of methodological expertise

The basic principles of clinical epidemiology and biostatistics are familiar to surgeons, but formal training in these specialties is rare. Without the appropriate amount of

methodological expertise, it has been difficult to transform the surgical culture to an evidence-seeking profession.⁶⁸ Research funding agencies have developed programmes to increase research exposure of junior members of the medical faculty, but there is some evidence that surgeons are less likely to apply for funding and are less successful when they do.⁶⁹ Perhaps as important for improved research is increased recognition of the need for collaboration between surgeons and methodologists, to enable high-quality and clinically relevant studies through the combination of expertise. Surgical and research communities and funding bodies need to recognise this gap in knowledge.

Academic careers and research support

Surgeons must devote a substantial proportion of their career development to their craft, irrespective of whether or not they choose an academic career. Surgical research with an RCT design is not favoured because of the protracted nature of these trials, which, when combined with the obligatory time commitments of the operating theatre, is currently not conducive for rapid career advancement.⁷⁰ By contrast, many established (funded) faculty development programmes are in place for basic science and medical disciplines in which pharmacological interventions are the predominant focus. Although some funding bodies have provided increasing support to improve surgical research, a disproportionately smaller number of surgeons are working in this specialty.⁷¹

Lack of regulation

Surgical research should generally follow the same ethical and scientific principles as pharmacological research. Worldwide mandatory regulations, such as the International Conference on Harmonisation guidelines, Directives of the European Union, and the US Food and Drugs Administration, have been developed for assessment of drugs. There are no regulatory procedures for licensing surgical treatments on the basis of high-quality evidence. However, this type of evaluation through assessment bodies has begun to appear in some developed countries.^{72,73} Unlike pharmacological evaluations, industry funding is limited and financing of research by health-care funding agencies is greatly needed. Whether a regulatory framework and an agency for surgical innovation would make a difference to the quality of surgical research is speculative.

Conclusions

Rigorous evaluation of new surgical interventions, although difficult, is achievable and necessary. The complexity of surgical procedures makes it difficult, if not generally impossible, to mirror some aspects of pharmacological research. This shortcoming has contributed to uncertainty about the risk of biases and has led to scepticism about the value of surgical research. Although much criticism is aimed at RCTs of surgical procedures, few of the challenges apply only to this type of study design; an RCT should be the default choice for an evaluation. A greater understanding of the processes of evaluation in surgery could lead to more high-quality studies. Surgery does not lack evaluative research. What it does not have are accepted guidelines for generating valid evidence: systematic, well-planned and conducted, and meticulously reported evidence, on which surgical practice can be based.

Acknowledgments

The Balliol Colloquium has been supported by Ethicon UK with unrestricted educational grants and by the National Institute of Health Research Health Technology Assessment Programme. The Balliol Colloquium was administratively and financially supported by the Nuffield Department of Surgery at the University of Oxford and the Department of Surgery at McGill University. JAC holds a Medical Research Council UK special training fellowship. The University of Aberdeen's Health Services Research Unit is core funded by the Chief Scientist

Office of the Scottish Government Health Directorates. IB is supported by a grant from the Société Française de Rhumatologie and Lavoisier Program (Ministère des Affaires Étrangères et Européennes). PLE is a DPhil Candidate in Evidence-Based Health Care at Oxford University.

References

1. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med*. 2008; 148:295–309. [PubMed: 18283207]
2. Taggart DP. CABG is still the best treatment for multivessel and left main disease, but patients need to know. *Ann Thorac Surg*. 2006; 82:1966–75. [PubMed: 17126093]
3. McCulloch P, Altman DG, Campbell WB, et al. for the Balliol Collaboration. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009; 374:1105–12. [PubMed: 19782876]
4. Solomon MJ, McLeod RS. Clinical studies in surgical journals—have we improved? *Dis Colon Rectum*. 1993; 36:43–48. [PubMed: 8416778]
5. Wente MN, Seiler CM, Uhl W, Buchler MW. Perspectives of evidence-based surgery. *Dig Surg*. 2003; 20:263–69. [PubMed: 12748428]
6. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008; 337:a2390. [PubMed: 19001484]
7. Freeman TB, Vawter DE, Leaverton PE, et al. Use of placebo surgery in controlled trials of a cellular-based therapy for Parkinson's disease. *N Engl J Med*. 1999; 341:988–92. [PubMed: 10498497]
8. Gordon B, Mackrodt C, Fern E, Truesdale A, Ayers S, Grant A. The Ipswich Childbirth Study: 1. wA randomised evaluation of two stage postpartum panned repair leaving the skin unsutured. *Br J Obstet Gynaecol*. 1998; 105:435–40. [PubMed: 9609272]
9. Neumayer L, Giobbie-Hurder A, Jonasson O, et al. Open mesh versus laparoscopic mesh repair of inguinal hernia. *N Engl J Med*. 2004; 350:1819–27. [PubMed: 15107485]
10. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA*. 2006; 296:2441–50. [PubMed: 17119140]
11. Yang SH, Zhang YC, Yang KH, et al. An evidence-based medicine review of lymphadenectomy extent for gastric cancer. *Am J Surg*. 2009; 197:246–51. [PubMed: 18722583]
12. Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast changing technologies: the case for tracker studies. *BMJ*. 2000; 320:43–46. [PubMed: 10617532]
13. Grant AM, Wileman SM, Ramsay CR, et al. Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ*. 2008; 337:a2664. [PubMed: 19074946]
14. Campbell, MK.; Torgerson, DJ. Wiley encyclopedia of clinical trials. John Wiley & Sons; Hoboken: 2007. Preference trials.
15. Taylor KM. The doctor's dilemma: physician participation in randomized clinical trials. *Cancer Treat Rep*. 1985; 69:1095–100. [PubMed: 4042089]
16. McCulloch P, Kaul A, Wagstaff GF, Wheatcroft J. Tolerance of uncertainty, extroversion, neuroticism and attitudes to randomized controlled trials among surgeons and physicians. *Br J Surg*. 2005; 92:1293–37. [PubMed: 16078288]
17. Harrison JD, Solomon MJ, Young JM, et al. Surgical and oncology trials for rectal cancer: who will participate? *Surgery*. 2007; 142:94–101. [PubMed: 17630005]
18. Margo CE. When is surgery research? Towards an operational definition of human research. *J Med Ethics*. 2001; 27:40–43. [PubMed: 11233377]
19. Mills N, Donovan JL, Smith M, Jacoby A, Neal DE, Hamdy FC. Perceptions of equipoise are crucial to trial participation: a qualitative study of men in the ProtecT study. *Control Clin Trials*. 2003; 24:272–82. [PubMed: 12757993]
20. Pocock, S. Methods of randomisation. John Wiley and Sons; New York: 1983. Clinical trials: a practical approach; p. 66-90.

21. Majeed AW, Troy G, Smythe A, et al. Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy. *Lancet*. 1996; 347:989–94. [PubMed: 8606612]
22. Higgins, JPT.; Altman, DG. Chapter 8. Assessing the risk of bias in included studies. In: Higgins, JPT.; Green, S., editors. *Cochrane handbook for systematic reviews of interventions version 5.0.1*. John Wiley & Sons; Chichester: 2008. (updated September 2008)
23. Boutron I, Guttet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med*. 2007; 4:e61. [PubMed: 17311468]
24. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2002; 347:81–88. [PubMed: 12110735]
25. London AJ, Kadane JB. Placebos that harm: sham surgery controls in clinical trials. *Stat Methods Med Res*. 2002; 11:413–27. [PubMed: 12357587]
26. Poolman RW, Struijs PA, Krips R, et al. Reporting of outcomes in orthopaedic randomized trials: does blind of outcome assessors matter? *J Bone Joint Surg*. 2007; 89:550–58. [PubMed: 17332104]
27. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ*. 2005; 330:88. [PubMed: 15637373]
28. Barker CF, Kaiser LR. Is surgical science dead? The Excelsior Society lecture. *J Am Coll Surg*. 2004; 198:1–19. [PubMed: 14698306]
29. Thabut G, Christie JD, Ravaud P, et al. Survival after bilateral versus single lung transplantation for patients with chronic obstructive pulmonary disease: a retrospective analysis of registry data. *Lancet*. 2008; 371:744–51. [PubMed: 18313503]
30. Hunt S. A fair way of donating hearts for transplantation. *BMJ*. 2000; 321:526. [PubMed: 10968801]
31. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2001; 334:349–51. [PubMed: 17303884]
32. Vandembroucke JP. Observational research, randomized trials, and two views of medical science. *PLoS Med*. 2008; 5:e67. [PubMed: 18336067]
33. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*. 1996; 312:1215–18. [PubMed: 8634569]
34. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess*. 1998; 2:1–124. [PubMed: 10103353]
35. Ioannidis JPA, Haidich A, Lau J. Any casualties in the clash of randomized and observational evidence? *BMJ*. 2001; 322:879–80. [PubMed: 11302887]
36. Weil RJ. The future of surgical research. *PLoS Med*. 2004; 1:e13. [PubMed: 15526041]
37. Reeves, BC.; Deeks, JJ.; Higgins, JPT.; Wells, GA. Chapter 13. Including non-randomized studies. In: Higgins, JPT.; Green, S., editors. *Cochrane handbook for systematic reviews of interventions version 5.0.1*. John Wiley & Sons; Chichester: 2008. (updated September 2008)
38. Edwards FH, Clar RE, Schwartz M. Practical considerations in the management of large multi-institutional databases. *Ann Thorac Surg*. 1994; 58:1841–44. [PubMed: 7979779]
39. Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin; Boston: 2002.
40. The EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. *N Engl J Med*. 1985; 313:1191–2000. [PubMed: 2865674]
41. North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med*. 1991; 325:445–53. [PubMed: 1852179]
42. National Emphysema Treatment Trial Research Group. Patients at high risk of death after lung-volume-reduction surgery. *N Engl J Med*. 2001; 345:1075–83. [PubMed: 11596586]

43. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000; 321:694–96. [PubMed: 10987780]
44. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008; 337:a1655. [PubMed: 18824488]
45. Fielding LP, Steward-Brown S, Dudley HAF. Surgeon-related variables and the clinical trial. *Lancet*. 1978; 2:778–79. [PubMed: 80697]
46. Bakaeen FG, Dhaliwal AS, Chu D, et al. Does the level of experience of residents affect outcomes of coronary artery bypass surgery? *Ann Thorac Surg*. 2009; 87:1127–33. [PubMed: 19324138]
47. Baskett RJ, Buth KJ, Légare JF, et al. Is it safe to train residents to perform cardiac surgery? *Ann Thorac Surg*. 2002; 74:1043–48. [PubMed: 12400743]
48. Edmunds, LH.; Cohn, LH., editors. Perioperative/intraoperative care. McGraw-Hill; New York: 2004. Cardiac surgery in the adult; p. 261-548.
49. Hawe P, Shiell A, Riley T. Complex interventions: how out of control can a randomised controlled trial be? *BMJ*. 2004; 328:1561–63. [PubMed: 15217878]
50. Hardoon SL, Lewsey JD, Gregg PJ, Reeves BC, van der Meulen JH. Continuous monitoring of the performance of hip prostheses. *J Bone Joint Surg Br*. 2006; 88:716–20. [PubMed: 16720761]
51. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. *Health Technol Assess*. 2001; 5:1–79. [PubMed: 11319991]
52. The Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. *N Engl J Med*. 1991; 324:1073–78. [PubMed: 1826143]
53. Vickers AJ, Bianco FJ, Serio AM, et al. The surgical learning curve for prostate cancer control after radical prostatectomy. *J Natl Cancer Inst*. 2007; 99:1171–77. [PubMed: 17652279]
54. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials*. 2009; 10:9. [PubMed: 19200379]
55. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clin Trials*. 2004; 1:421–27. [PubMed: 16279280]
56. Cook JA, Ramsay CR, Fayers P. Using the literature to quantify the learning curve. *Int J Technol Assess Health Care*. 2007; 23:255–60. [PubMed: 17493312]
57. Birkmeyer JD, Dimick JB, Birkmeyer NJ. Measuring the quality of surgical care: structure, process, or outcomes? *J Am Coll Surg*. 2004; 198:626–32. [PubMed: 15051016]
58. Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. *Health Technol Assess*. 2001; 5:1–194. [PubMed: 11532239]
59. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004; 240:205–13. [PubMed: 15273542]
60. Clavien PA, Barkun J, DeOliveira ML, et al. The Clavien-Dindo classification of surgical complications. Five-year experience. *Ann Surg*. (in press).
61. DeOliveira ML, Winter JM, Schafer M, et al. Assessment of complications after pancreatic surgery: a novel grading system applied to 633 patients undergoing pancreaticoduodenectomy. *Ann Surg*. 2006; 244:931–39. [PubMed: 17122618]
62. Blazeby JM, Avery K, Sprangers M, Pikhart H, Fayers P, Donovan J. Health-related quality of life measurement in randomized clinical trials in surgical oncology. *J Clin Oncol*. 2006; 24:3178–86. [PubMed: 16809741]
63. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials*. 2007; 8:38. [PubMed: 18039364]
64. Davis D, Evans M, Jadad A, et al. The case for knowledge translation: shortening the journey from evidence to effect. *BMJ*. 2003; 327:33–35. [PubMed: 12842955]
65. Wright FC, Law CH, Last LD, Klar N, Ryan DP, Smith AJ. A blended knowledge translation initiative to improve colorectal cancer staging [ISRCTN56824239]. *BMC Health Serv Res*. 2006; 6:4. [PubMed: 16412251]

66. Meakins JL. Evidence-based practice: new techniques and technology. *Can J Surg.* 2001; 44:84–85. [PubMed: 11308239]
67. Oxman A, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. *JAMA.* 1993; 270:2093–95. [PubMed: 8411577]
68. Rothenberger DA. Evidence-based practice requires evidence. *Br J Surg.* 2004; 91:1387–88. [PubMed: 15499647]
69. Rangel SJ, Moss RL. Recent trends in the funding and utilization of NIH career development awards by surgical faculty. *Surgery.* 2004; 136:232–39. [PubMed: 15300185]
70. Fischer L, Bruckner T, Diener MK, et al. Four years of teaching principles in clinical trials—a continuous evaluation of the postgraduate workshop for surgical investigators at the study center of the German Surgical Society. *J Surg Educ.* 2009; 66:15–19. [PubMed: 19215892]
71. Rahbari NN, Diener MK, Fischer L, et al. A concept for trial institutions focussing on randomised controlled trials in surgery. *Trials.* 2008; 9:3. [PubMed: 18218081]
72. Maddern G, Boulton M, Ahern E, Babidge W. ASERNIP-S: International trend setting. *ANZ J Surg.* 2008; 78:853–58. [PubMed: 18959636]
73. Plumb J, Campbell B, Lyratzopoulos G. How guidance on the use of interventional procedures is produced in different countries: an international survey. *Int J Technol Assess Health Care.* 2009; 25:124–33. [PubMed: 19366495]

Figure.
Complexity of a surgical intervention

Table

Challenges to the planning and conduct of randomised trials comparing a surgical intervention with different types of comparator

	Sham (placebo) surgery (eg, stem cells for Parkinson's disease ⁷)	Other similar surgical procedure (eg, two-stage vs three-stage postpartum perineal repair ⁸)	Substantially different surgical procedures (eg, open vs minimally invasive procedures ⁹)	Non-surgical treatment (eg, back surgery vs physiotherapy ¹⁰)
Patient reluctance to participate	Yes	Unlikely	Likely	Yes
Randomisation in operating theatre possible	Yes	Yes	Likely	No, because of different health-care providers
Imbalance in surgical expertise	No	Unlikely	Likely	No, because of different health-care providers
Poor compliance with allocation (ie, crossover)	Yes	Unlikely	Yes	Yes
Contamination (ie, lack of fidelity)	Unlikely	Yes	Unlikely	No