This article was originally published in The Lancet 2009;374(9695):1097-1104 doi:10.1016/S0140-6736(09)61086-2 and is available from URL: http://www.sciencedirect.com/science/journal/01406736

TITLE:

Challenges In Evaluating Surgical Innovation

AUTHORS:

Patrick L. Ergina, MD, MPH, MSc, FRCSC, FACS

McGill University, Balliol College - Oxford University

Jonathan A. Cook, PhD

Health Services Research Unit, University of Aberdeen

Jane M. Blazeby, MD, MSc, FRCS

University of Bristol

Isabelle Boutron, MD, PhD

University of Paris VII

Pierre-Alain Clavien, MD, PhD, FRCS, FACS

University Hospital of Zurich

Barnaby Reeves, MSc, DPhil

University of Bristol

Christoph M. Seiler, MD, MSc

University Hospital of Heidelberg

And the Balliol Collaboration

Authors Contribution Statement.

This paper was written by the **writing group comprised** of: **P.L. Ergina, J.A. Cook[i], J.M. Blazeby, I. Boutron, P.A. Clavien, B. Reeves, C.M. Seiler** and the **Balliol Collaboration**. The group agreed to the form and content of the paper at a meeting on 3rd April 2009. P. Ergina and J. Cook 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.

The **Balliol Collaboration** comprises the clinicians and methodologists who met at Balliol College, Oxford University, on April 1-3, 2009. This group agreed to an outline, the form, and content of all three papers in this series, and subsequently formed three subgroups to write each paper: They were:

Doug	Altman
Jeffrey	Aronson
Jeffrey	Barkun
Jane	Blazeby
Isabelle	Boutron
Bruce	Campbell
Marion	Campbell
Pierre-Alain	Clavien
Jonathan	Cook

[i] Jonathan Cook holds a Medical Research Council UK special training fellowship. The University of Aberdeen's Health Services Research Unit is funded by the Scottish Government Health Directorates.

Patrick	Ergina
Liane	Feldman
David	Flum
Paul	Glasziou
Guy	Maddern
John	Marshall
Peter	McCulloch
Jonathan	Meakins
Jonathan	Nicholl
Barney	Reeves
Christoph	Seiler
Jan	Vandenbroucke

The Balliol Collaboration were a writing group derived from the **Balliol Colloquium**, a group of clinicians and methodologists who took part in three conferences at Balliol College, Oxford University, on the topic of *Surgical Innovation and Evaluation* organized by **Professor Jonathan L. Meakins**, between 2007-2009. They were:

Doug	Altman
Jeffrey	Aronson
Deborah	Ashby
Jeffrey	Barkun
Nick	Black

Jane	Blazeby
Isabelle	Boutron
John	Bunker
Martin	Burton
Bruce	Campbell
Marion	Campbell
Kalipso	Chalkidou
Iain	Chalmers
Pierre-Alain	Clavien
Jonathan	Cook
Marc	de Leval
Jon	Deeks
Patrick	Ergina
Liane	Feldman
David	Flum
Paul	Glasziou
Adrian	Grant
Muir	Gray
Roger	Greenhalgh
Milos	Jenicek
Sean	Kehoe
Richard	Lilford

Peter	Littlejohns
Yoon	Loke
Guy	Maddern
Rajan	Madhock
John	Marshall
Peter	McCulloch
Kim	McPherson
Jonathan	Meakins
Jonathan	Nicholl
Barney	Reeves
Peter	Rothwell
Christoph	Seiler
Steven	Strasberg
Bill	Summerskill
David	Taggart
Parris	Tekkis
Matthew	Thompson
Tom	Treasure
Ulrich	Trohler
Jan	Vandenbroucke

Conflict of Interest Statement:

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. None of the members has any conflict of interest to declare.

Word Count:

Body of paper: 4004

Abstract: 149

ABSTRACT:

This paper is the second of three from the Balliol Collaboration on *evaluating surgical innovation*, and considers the *methodological and practical challenges* to rigorous surgical research. Few, if any, of these challenges apply only to surgery. However, what is arguably unique to surgery is the way in which many of these challenges coalesce. We discuss the impact of the challenges confronting surgical evaluations in two areas: *(1) challenges related to study design*; and *(2) challenges related to the nature of surgical interventions.* The first is divided into: challenges of randomised controlled trials for evaluating surgical interventions, and challenges of nonrandomised studies for evaluating surgical interventions. The second: complexity and surgical interventions; surgeon-related factors; surgical outcome evaluation; and additional challenges in surgery. We conclude that rigorous evaluation, though difficult, is achievable and necessary. It requires surgical "flavoured" solutions and a framework for generating evidence on which to base surgical practice.

I. INTRODUCTION

The *Balliol Colloquia* considered many challenges facing the surgical research community when performing an evaluation of a therapeutic, procedure-based intervention. By identifying many issues and deconstructing them into constituent methodological parts, several important areas were targeted for developing a systematic process that would guide appropriate, *evidence-based surgical practice*. This paper is the second of three on evaluating surgical innovation, and considers the methodological and practical challenges to rigorous surgical research[i]. Few, if any, of these challenges apply only to surgery. Many arise when evaluating 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 coalesce. Perhaps this leads many surgeons to view Randomised Controlled Trials (RCTs), while theoretically advantageous, to be too difficult and impractical to conduct, and at worst, irrelevant to their practice due to concerns about generalizability.[2] Clearly, most of the same challenges also affect non-randomised studies and in some cases to a greater extent. In spite of the barriers, an RCT remains the optimal study design for evaluating therapeutic interventions. We shall discuss the impact of the challenges confronting surgical evaluations in two broad areas:

(1) challenges related to study design; and

(2) challenges related to the nature of surgical interventions.

[i] Recommendations for improvement and solutions will be presented in the third paper in this series.

II. CHALLENGES RELATED TO STUDY DESIGN

A. CHALLENGES OF RANDOMISED CONTROLLED TRIALS (RCTs) FOR EVALUATING SURGICAL INTERVENTIONS

Randomised controlled trials (RCTs) are usually considered the gold standard for evaluating therapy. Despite exhortations for surgical research to be more rigorous, the overall prevalence of RCTs has been consistently low since the 1970s.[3] Large, high-quality RCTs have been conducted 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 pre and post-operative care.[4] It should be noted that there are different types of RCTs which address different aims[5] 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. The latter are required to ensure surgical practice is evidence-based. Some criticisms of surgical trials are misplaced and reflect misunderstanding of the trials' aims. Specific challenges to the planning and conduct of surgical randomised trials are discussed below and summarised in Table 1. Many of the issues raised may require a different approach depending upon the aim of the study.

(i) When Should An RCT Be Conducted?

A particularly vexing question in evaluating new surgical interventions is whether an RCT is necessary and, if so, when the first one should be conducted. Conceptually, there are few arguments against doing RCTs early in development, though a further evaluation may 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 evaluation. If too early, the constraints of an RCT could obstruct innovation, and if too late, *equipoise* may be lost. Another consequence of an early RCT is that the definitive technique may not be fully refined; the subsequent study outcome then reflects the stage of development and learning, not the therapeutic impact of the intervention.[6] In addition, restricting a new procedure to an RCT may be impractical in the absence of regulation.[7]

(ii) Difficulties In Recruitment

When two interventions have different benefit to harm profiles, patients and surgeons may strongly prefer one intervention. *Strong preferences* represent a lack of equipoise, which may lead both patients and surgeons to decline participation and make trial recruitment more difficult if not infeasible. This is particularly problematic when the preferred intervention is already available. The strength of a patient's preference partly depends on the comparison of interest (Table 1). Possible comparisons are: a new procedure vs. a "sham" (placebo) procedure; similar but distinguishable procedures; dramatically different procedures; or surgery vs. a non-surgical treatment such as a medical treatment,

participative intervention (e.g. rehabilitation), or “watchful waiting”. Research to evaluate surgical innovations in emergency and pediatric settings are perhaps particularly susceptible to preferences which preclude randomisation.

Response to uncertainty has been also suggested as an explanation why surgical trials may be more difficult to randomise. When compared to physicians, surgeons may be less tolerant of uncertainty, affecting their participation in RCTs.[8] Previous negative experiences and perceived threat of litigation may make some surgeons reluctant to submit elements of their practice to evaluation. A feasibility study of patients’ willingness to participate in surgical and oncology trials found a low level of willingness due to a stated dislike for randomisation, and a desire to make their own decisions about the selection of the intervention.[9] Preferences, both surgeon and patient, may often not be based on the existing evidence. Patients need to be provided with sufficient information to make informed decisions which has not always have been the case.[10] Q*ualitative research* can provide insights into the recruitment process and may enable greater RCT participation.[11]

(iii) Timing Of The Randomisation Process And Adherence To Allocation

Randomisation should be as close to the time of the intervention as possible to reduce the possibility that the allocated intervention will not be delivered due to strong preferences, knowledge of allocation assignment, cancellations, or clinical events before the procedure.[12] However, it needs to be sufficiently early for the

patient and surgical team to be adequately prepared. In the case of two possible surgical procedures, randomisation can often be done in the Operating Room (OR). The challenge is exaggerated when comparing substantially different interventions (e.g. surgical vs. pharmacological), where participants have to be told their allocation in advance of receiving the intervention, or if the new procedure is available outside the trial. Irrespective of the timing of randomisation, a surgeon may decide that a surgical procedure is inappropriate, impossible, or unsafe after randomization. The SPORT trial, which compared surgery and non-surgical management in low back pain, observed considerable *crossing over*.[13] Although the principle of *intention-to-treat* provides the preferred analysis framework for dealing with this, applying the results of such an analysis to other settings (e.g., if cross-overs are predominantly from new to conventional therapy) can be problematic.

(iv) Lack Of Blinding

Lack of blinding can lead to several forms of bias: *performance* bias (surgeons, other caregivers or patients choosing co-interventions conditional on their allocation); *attrition* bias (differential withdrawal of from follow-up); and *detection* bias (differential outcome assessment).[14] Blinding of surgeons, patients, and other caregivers is difficult and often impossible in surgical trials; nevertheless, innovative methods of blinding exist.[15] In a comparison of laparoscopic and small-incision cholecystectomy, bloody bandages were used to blind patients and other caregivers.[16] "*Sham" (placebo) surgery*, whereby the surgeon mimics the

intervention, have been used to assess arthroscopy and stem cells for Parkinson's disease.[17] [18] Their use is controversial, and they have been restricted to cases where a suitable comparator was not available or the *placebo* surgery had limited risk.[19] Although blinding of the surgeon and patient is difficult, it should be possible to blind the clinical assessment of outcomes (though seldom done).[20] If patients cannot be blinded, some outcomes may be susceptible to bias (especially *patient reported outcomes*).

(v) Alternative RCT Designs

The principle of randomising participants to *surgeons with expertise* in carrying out different procedures, i.e. an *"expertise-based" design* (an intrinsic feature of a comparison between a surgical procedure and a non-surgical treatment), has been proposed for comparing two surgical procedures.[21] Like cluster randomisation, this design protects against contamination and allows surgeons with strong preferences to take part. However, it brings its own challenges: more surgeons are required; the comparison may 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 evaluation.[22] In this design, modifications during the conduct of the trial are allowed, recorded, and subsequently "tracked" in the statistical analysis. Variations in the randomisation

scheme are also allowed. In principle, the full development of an innovation can be assessed in a single study. While conceptually attractive, full-fledged tracker trials would be extremely challenging in practice.

B. CHALLENGES OF NONRANDOMISED STUDIES (NRS) FOR EVALUATING SURGICAL INTERVENTIONS

As noted above, various factors contribute to make RCTs of surgical procedures difficult and, in a few cases, impossible. For example, lesser surgical variations may have such limited influence on outcome so as to make a RCT evaluation prohibitively large. Historically, most advances in surgical knowledge have been accepted on the basis of *non-randomised studies (NRS)*.[23] Surgical interventions such as heart, liver, kidney, and lung transplantation are established therapies in developed countries.[24] None of these have been validated with RCTs, and it is generally considered unethical to do so given the apparent benefit.[25] Other advances have been identified through observational studies, or even anecdotes, because of "dramatic effects", where bias may be less concerning.[26] Exploratory initial *case reports*, or early *case series*, for new procedures are likely to be reported as an NRS.[27] 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 may not have been looked for or detected in original studies; and to *estimate treatment effects* when RCTs were deemed infeasible (e.g., rare events, observations far in the future).[28] Where RCTs are not possible it is

essential to conduct high quality nonrandomized studies.[29]. A dichotomy between randomized and non-randomised studies is somewhat artificial as both designs can provide different and complementary evidence.[30] For example, a non-randomised evaluation of longer term and rare safety outcomes could be conducted alongside a RCT. Overall, most surgical studies are non-randomized and often retrospective; their quality is also extremely variable and often poor.[31] Prospective comparative designs are substantially more useful case series which are over represented. A significant driving factor behind non-randomized studies is that they are easier to perform, increasingly so with electronic data collection and standardized databases.[32] 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 problems that tend to undermine the validity of non-randomised studies.

Protocol driven studies, which account for all cases and have accurate and informative clinical data, are needed. More attention should be focused on data collection to reduce bias due to incomplete data as is the case for RCTs. However, even well designed NRS suffer from many of the difficulties faced in conducting RCTs.[33] For example, the existence of a learning curve (see Section IIIB) is a challenge to both *randomised* and *non-randomised* evaluations. Accounting for any pre-treatment differences between intervention groups is a particular concern in non-randomized studies. Rigorous prospective design, and data collection, provide some protection but there is still an underlying reliance

on clinical understanding of the condition and its risk factors, along with appropriate documentation for any statistical adjustment (such as propensity scores) to have validity.[34]

Finally, *causal inferences* established in a NRS are considered "weaker" than an RCT and needs cautious interpretation.[35] Cautionary examples of "established" surgical practices validated with NRS and subsequently discontinued after a large RCT was conducted abound (e.g. EC-IC bypass[36], Carotid Endarterectomy widespread use[37], lung volume-reduction surgery[38]). Since current advances have been more subtle, the need for RCTs should increase, i.e., the smaller the difference in outcome, the greater the need for an RCT.

III. CHALLENGES RELATED TO THE NATURE OF SURGICAL INTERVENTIONS

A. COMPLEXITY AND SURGICAL INTERVENTIONS

We need to recognize that many surgical interventions are *complex* and require appropriate evaluation.[39] Surgical interventions, like other *non-pharmacological interventions* such as therapist-based and educational interventions, consist of a number of components that cannot be separated.[40] This contrasts with most *pharmacological interventions,* which can be readily defined and standardised. While the *surgical procedure* requires attention — a *surgical intervention* may depend on many healthcare professionals and involve other aspects of healthcare delivery in ways that a pharmacological intervention does not.

(Figure 1)

A *surgical procedure* is primarily delivered by a *surgeon*, and is influenced by *characteristics* such as surgical skill, decision-making, preferences, and experience. The delivery of a *surgical intervention* additionally depends on the other members of the team (e.g., anaesthetists, nurses, technicians) and pre-operative and post-operative management (e.g., emergency department, imaging services, recovery room, intensive care, and rehabilitation programmes). This complexity often receives limited recognition in the design of surgical evaluations.

Indeed, its existence is sometimes used to criticise surgical evaluations for failing to control for potential *confounding factors*.[41]

A typical complex surgical intervention consists of several interacting components. Consider *Coronary Artery Bypass Graft surgery (CABG)*. Its *aim* is to revascularize the myocardium by bypassing coronary arteries that are stenosed or blocked. This is achieved by a number of steps that together constitute the *surgical procedure*: opening the chest; harvesting conduits; attaching (and later separating) the heart-lung machine; performing the anastomoses; reanimating the heart; closing the chest. In the case of CABG, there is limited variation in technique in between surgeons.[42] [43] However, there are many recognized variations in *surgical strategy*, such as "Off-Pump" CABG ("OPCAB", avoidance of the heart-lung machine), minimally invasive approaches ("MIDCAB"), and different choices of bypass conduits (e.g., bilateral mammary arteries, radial arteries, etc.). Some decisions are made intra-operatively (e.g., whether additional grafts are required) and will depend on the judgment of the individual surgeon. Other *co-interventions* may be used, such as antifibrinolytic agents, insulin, or hypothermia. *Pre-operative medical care* (e.g., Coronary Care Unit-cardiology management, medical management of co-morbidities, blood bank management, etc.), roles of other members of the *surgical team* (e.g. nurses, anaesthetists, perfusionists) and *post-operative care* (e.g. intensive care, acute and chronic cardiac rehabilitation) also vary and affect outcomes.[44] These

supporting components vary between centres, influenced by infrastructure, staffing, and local policies.

While an intervention needs to have a coherent *aim* (or "*function*"), different *forms* are often available.[45] 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 evaluating? Secondly, when evaluating alternatives, how standardised should they be, given the complexity of the steps involved? Continuing the CABG example, does avoidance of the heart-lung machine warrant evaluation (OPCAB)? If so, how standardised should the OPCAB surgical strategy and other steps be? The impact on health services (e.g., 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 evaluation 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 process information and monitoring required will depend on how an intervention is defined and the degree of standardisation sought. Very restrictive approaches could limit surgeon participation and may not be feasible in some centres.

B. SURGEON-RELATED FACTORS

As noted earlier, attributes of the surgeon, such as *surgical knowledge*, *prior training and experience*, and *inherent skills*, will influence the delivery of a surgical intervention and lead to variability in practice and health outcomes. Variability may be expected irrespective of prior 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. Evaluation in a realistic setting is important.

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 evaluation tend to focus on complications, and this where most of the learning curve literature resides.[46] [47] For example, the rate of bile duct injuries associated with laparoscopic cholecystectomy fell as the surgeon's experience increased. Proxies for operative expertise, such as duration and blood loss, have also been used.[48] The impact on health outcomes is subject to debate and likely to vary between interventions. For complex operations (e.g., radical prostatectomy and laparoscopic hernia repair) learning can continue over a very long time, perhaps hundreds of procedures.[49] [50]

Evaluation of a new surgical intervention vs. an established control has been criticised, owing to a perceived imbalance of experience, favouring the

established comparator.[51] Some have sought to conduct an evaluation with 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.[52] Trial design could be modified to incorporate individual learning. Evaluations of surgical innovations should consider the influence 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.[53]

C. SURGICAL *OUTCOME* EVALUATION

The key questions with regard to surgical evaluation are: *"what"* (what is the outcome), *"how"* (how to measure), *"who"* (who should assess) and *"when"* (when to assess)? The *Quality Assurance* literature has used the terms "structure", "process", and "outcomes" as suggested aspects for evaluation.[54] Traditionally, surgeons themselves have selected and assessed the outcomes, mainly focusing on short-term clinical measures of technical success and harm, although such outcomes have often not been standardised and reproducible, hindering evaluation. For example, a systematic review revealed 56 separate definitions of "anastomotic leak" at any site after gastrointestinal surgery, precluding comparability.[55] The lack of standardised (agreed upon) surgical terminology for the definition of clinical outcomes has long been recognised and

this has led to methods of grading and classifying deviations from the normal postoperative course have been developed, tested, modified, and validated.[56] One proposed strategy to is to use a validated "therapy-oriented" *complication classification system*, which ranks negative events by severity and avoiding confusing terms (TABLE 2).[57] It could be adjusted to match a clear and consensus definition of specific postoperative events within specific fields of surgery.[58]

Although these surgeon-selected *("physician-centred"*) clinical outcomes (e.g. mortality and morbidity rates) are critically 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 may differ from the surgeon's assessment and patients may value different outcomes to those of interest to surgeons (e.g. social, emotional function). Therefore, surgical evaluations require assessment of clinical *and* patient reported outcomes (PROs), i.e., patients, self-reporting without interpretation by an observer. Typically this information is captured in questionnaires assessing health-related quality of life (HRQL). It can be difficult to decide which are best suited to a particular problem. Methodology to select and incorporate assessment of HRQL into trials is emerging and better-performed studies will produce more reliable data. Despite the recent interest in this area, however, there appears to be a gap between measuring HRQL outcomes and using the information to influence surgical practice.[59] This may occur because the surgical community does not

understand HRQL data, or because clinical outcomes are considered paramount. Methods to accurately measure and interpret PROs alongside clinical data are needed so that surgeons can effectively evaluate surgery from these different perspectives and correspondingly inform patients. There are also situations where PROs matter more than clinical outcomes (e.g. palliative surgery, functional outcomes of joint replacement surgery) and capturing this data within well-designed trials is critical to optimise outcomes of importance to patients. Core *outcomes* will include key clinical, technical, and patient-reported outcomes (generic and/or disease-specific). Additionally, economic evaluation is critical for efficient use of often limited resources. Methods to reach agreement about these have been developed in rheumatology.[60]

What is needed is a more comprehensive approach to evaluating surgery using accurate standardized clinical and patient-reported outcomes, recorded in real time, and whenever possible by an independent blinded observer. After the early development of surgical interventions, comprehensive evaluation of outcomes is recommended for all other stages of development (see table paper 3). This approach provides information to allow evidence-based comparisons among different interventions.

D. ADDITIONAL CHALLENGES IN SURGERY

(i) Traditional Master-Student Model

The traditional hierarchal system of surgery epitomizes *"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. It can constrain penetration of new models and information into mainstay practice. Despite attempts to implement change with "knowledge translation" methods, dissemination of practice guidelines is poor without the assistance of "opinion leaders" in surgery.[61] This may help explain, in part, the slow acceptance of "*Evidence Based Surgery*" and RCTs in particular. Meakins' editorial introducing the first *"Users' Guide for Evidence Based Surgery"* did not appear until 2001,[62] well after the introduction of "*Users' Guides To The Medical Literature*" in 1993.[63]

(ii) Lack Of Methodological Expertise

The basic principles of *clinical epidemiology and biostatistics* are familiar to surgeons, but formal training is rare. Without a critical mass of methodological expertise it has been difficult to transform the surgical culture to an *"evidence seeking profession".*[64] Research funding agencies have developed programmes to increase research exposure of junior medical faculty, but there is some evidence that surgeons are less likely to apply and are less successful when they do.[65] 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 evaluations by combining expertise. The surgical

and research communities as well as funding bodies need to recognize this gap in knowledge.

(iii) Academic Careers And Research Support

Surgeons must devote a significant proportion of their career development to their "craft" regardless of a choice for an academic pathway. Surgical research using RCT design methods are not favoured due to their protracted nature and when combined with the obligatory time commitments of the operating room, are currently not conducive for rapid career advancement.[66] In contrast, many established (funded) faculty development programmes are in place for basic science and medical disciplines when pharmacological interventions are the predominant focus. This has not been the case for surgical researchers, although some funding bodies have provided increasing support to improve this, a disproportionately smaller number of surgeons are working in this area.[67]

(iv) Lack Of Regulation

Surgical research should generally follow the same ethical and scientific principles as pharmacological research. Worldwide *mandatory regulations,* like ICH guidelines, Directives of the European Union and the U.S. FDA, have been developed to evaluate drugs. In surgery, no parallel exist requiring high quality evidence prerequisite for full adoption. Fortunately, this type of evaluation through assessment bodies has begun to appear in some developed countries.[68] [69] Unlike pharmacological evaluations, industry funding is limited and financing

of research by healthcare funders is particularly needed. Whether a regulatory framework and an agency for surgical innovation would make a difference to the quality of surgical research is speculative.

IV. DISCUSSION

Rigorous evaluation of surgical innovation, though difficult, is achievable and necessary. There are many challenges, perhaps none of which is unique to surgery. However, they often converge in evaluation of new surgical procedures and require surgically “flavoured” solutions. The complexity of surgical interventions makes it difficult, if not generally impossible, to mirror some aspects of pharmacological evaluations. This has contributed to uncertainty about the risk of biases and has lead to scepticism about the value of surgical evaluations. In some cases, it has become too easy to criticise and ignore if it contradicts cherished beliefs or preferences. While much criticism is aimed at surgical RCTs, few of the challenges apply only to RCTs. It is difficult to predict which research questions can be evaluated using a RCT design given appropriate funding. The RCT design should be the default choice for a definite evaluation. Greater understanding of the processes of evaluation in surgery may lead to more high-quality studies. Surgery does not lack “evaluative” research. What it does lack are accepted guidelines for generating valid evidence: systematic, well planned and conducted, and meticulously reported evidence, on which surgical practice can be based.

TABLE 1: RCT CHALLENGES BY COMPARATOR

CHALLENGE	COMPARATOR			
	Sham (placebo) surgery	Other similar surgical procedure	Substantially different surgical procedures	Non surgical treatment
	• Stem cells for Parkinson's (Freeman[18])	• Two versus three stage postpartum perineal repair (Gordon [70])	• Open vs. minimally invasive procedures (Neumayer[50])	• Back surgery vs. physiotherapy (Weinstein[13])
Randomisation in operating room possible	Yes	Yes	Likely	No, due to different providers
Poor patient participation	Yes	Unlikely	Unlikely	Yes
Imbalance in surgical expertise	No	Unlikely	Yes	No, due to different providers
Poor compliance with allocation (i.e. cross over)	Yes	Unlikely	Yes	Yes
Contamination (i.e. lack of fidelity)	Unlikely	Yes	Unlikely	Unlikely

TABLE 2

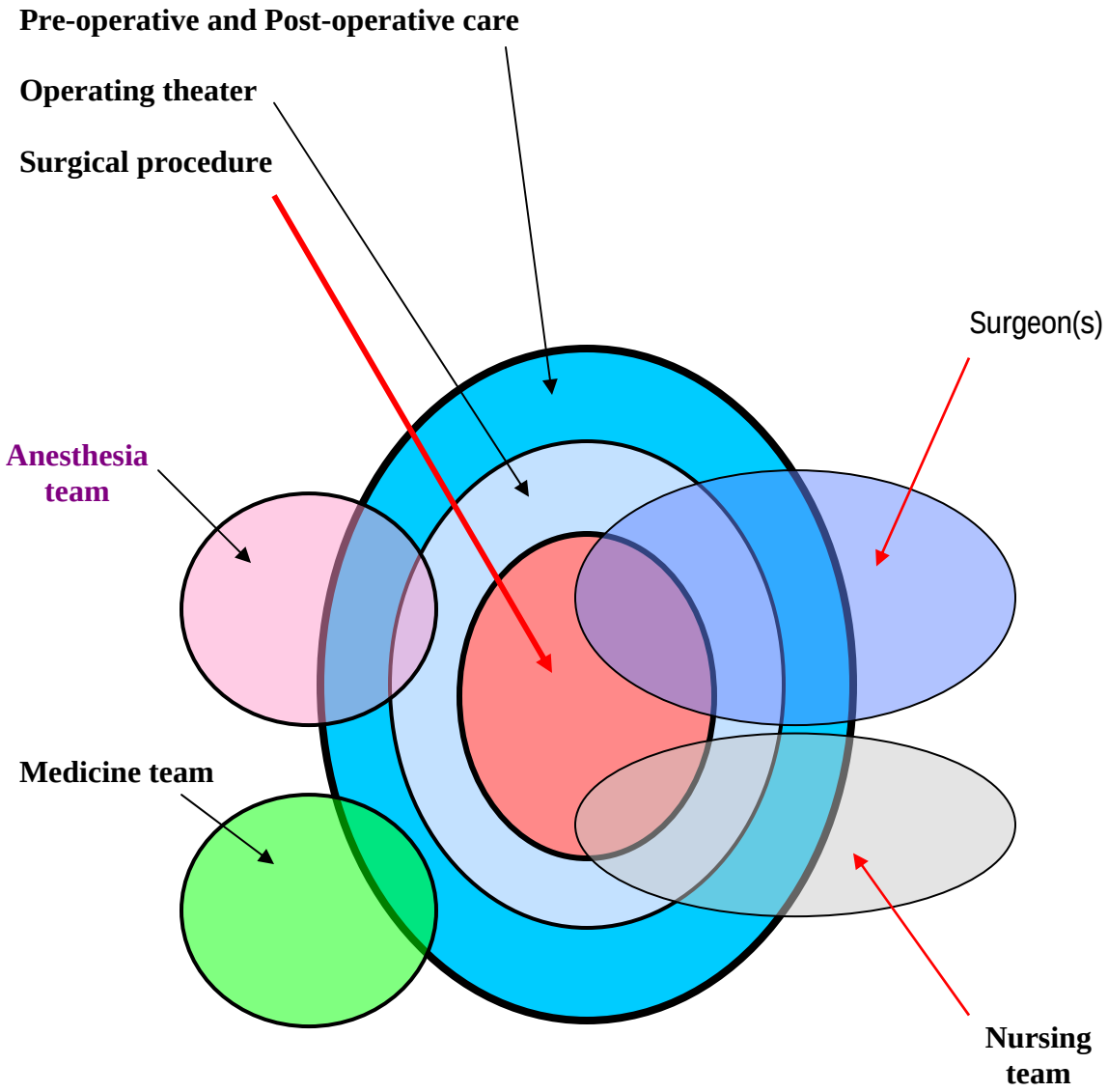
Classification of surgical complications

Grade	Definition
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, analgetics, 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 also included.
Grade III	Requiring surgical, endoscopic or radiological intervention
Grade IIIa	Intervention not under general anesthesia
Grade IIIb	Intervention under general anesthesia
Grade IV	Life-threatening complication (including CNS complications)[a] requiring IC/ICU management
Grade IVa	Single organ dysfunction (including dialysis)
Grade IVb	Multiorgan dysfunction
Grade V	Death of a patient

Grading system proposed in 2004.[i] 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.

[i] 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: 205–213.

FIGURE 1 – COMPLEXITY OF A SURGICAL INTERVENTION



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(4): 295-309.

[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.

[3] Solomon MJ, McLeod RS. Clinical studies in surgical journals-have we improved? Dis Colon Rectum 1993;36(1):43-8.

[4] Wente MN, Seiler, CM, Uhl W, Buchler MW. Perspectives of evidence-based surgery. Dig Surg 2003;20(4):263-9.

[5] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D: Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008, 337:a2390.

[6] Yang SH, Zhang YC, Yang KH, Li YP, He XD, Tian JH, Lv TH, Hui YH, Sharma N. An evidence-based medicine review of lymphadenectomy extent for gastric cancer. Am J Surg 2009 Feb;197(2):246-51.

[7] Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast changing technologies: the case for tracker studies. BMJ 2000; 320: 43-46

[8] 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 Oct;92(10):1293-7.

[9] Harrison JD, Solomon MJ, Young JM, Meagher A, Hruby G, Salkeld G, Clarke S. Surgical and oncology trials for rectal cancer: who will participate? Surgery.2007;142(1):94-101.

[10] Margo CE. When is surgery research? Towards an operational definition of human research. Journal of Medical Ethics 2001;27:40–43.

[11] 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 Jun;24(3):272-82.

[12] Pocock S. Clinical Trials: A Practical Approach. Chapter 5: Methods of Randomisation. 1983. New York: John Wiley and Sons. Pp. 66-90.

[13] Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, Abdu WA, Hilibrand AS, Boden SD, Deyo RA. Surgical vs nonoperative

treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. JAMA 2006, 296:2441-2450.

[14] 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. The Cochrane Collaboration, 2008.

[15] Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. PLoS Med 2007;4(2):e61

[16] Majeed AW, Troy G, Nicholl JP, Smythe A, Reed MW, Stoddard CJ, Peacock J, Johnson AG. Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy. Lancet 1996;347:989-994.

[17] Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 2002;347:81-8.

[18] 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(13):988-92.

[19] London AJ, Kadane JB. Placebos that harm: sham surgery controls in clinical trials. Stat Methods Med Res 2002;11:413-27.

[20] Poolman RW, Struijs PA, Krips R, Sierevelt IN, Marti RK, Farrokhyar F, Bhandari M. Reporting of outcomes in orthopaedic randomized trials: does blind of outcome assessors matter? J Bone Joint Surg 2007;89:550-8.

[21] Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, Sackett DL, Cinà CS, Walter SD, Haynes B, Schünemann HJ, Norman GR, Guyatt GH. Need for expertise based randomised controlled trials. BMJ 2005;330:88.

[23] Barker CF, Kaiser LR. Is surgical science dead? The Excelsior Society lecture. J Am Coll Surg 2004;198(1):1-19.

[24] 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

[25] Hunt S. A fair way of donating hearts for transplantation. BMJ 2000; 321:526.

[26] Glasziou P, Chalmers I, Rawlins M, McCulloch P When are randomised trials unnecessary? Picking signal from noise. BMJ 2001;334(7589):349-51.

[27] Vandenbroucke JP. Observational research, randomized trials, and two views of medical science. PloS Med 2008;5(3):e67.

[28] Black N. Why we need observational studies to evaluate the effectiveness of health care BMJ. 1996;312(7040):1215-8

[29] Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomized and non-randomised studies: a systematic review. HTA 1998;Vol.2:No.13.

[30] Ioannidis JPA, Haidich A, Lau J. Any casualties in the clash of randomized and observational evidence? BMJ 2001;322:879-80.

[31] Solomon MJ, McLeod RS. Clinical studies in surgical journals-have we improved? Dis Colon Rectum 1993;36(1):43-8

[32] Weil RJ. The future of surgical research PLoS Med. 2004 Oct;1(1):e13

[33] Reeves BC, Deeks JJ, Higgens 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. 2008. Mississauga: John Wiley & Sons.

[34] Edwards FH, Clar RE, Schwartz M. Practical considerations in the management of large multi-institutional databases. Ann Thorac Surg 1994;58 : 1841-44.

[35] Shadish WR, Cook TD, Campbell DT. Experimental And Quasi-experimental Designs For Generalized Causal Inference. 2002. Boston: Houghton Mifflin.

[36] 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(19):1191-2000.

[37] North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. N Engl J Med 325 (1991), pp. 445–453.

[38] National Emphysema Treatment Trial Research Group. Patients at high risk of death after lung-volume-reduction surgery. N Engl J Med 2001;345:1075–1083.

[39] Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. BMJ. 2000 Sep 16;321(7262):694-6.

[40] 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.

[41] Fielding LP, Steward-Brown S, Dudley HAF. Surgeon-related variables and the clinical trial. Lancet 1978; 2(8093):778-9.

[42] Bakaeen FG, Dhaliwal AS, Chu D, Bozkurt B, Tsai P, LeMaire SA, Wall MJ Jr, Coselli JS, Huh J. Does the level of experience of residents affect outcomes of coronary artery bypass surgery? Ann Thorac Surg. 2009 Apr;87(4):1127-33; discussion 1133-4.

[43] Baskett RJ, Buth KJ, Legaré JF, Hassan A, Hancock Friesen C, Hirsch GM, Ross DB, Sullivan JA. Is it safe to train residents to perform cardiac surgery? Ann Thorac Surg. 2002 Oct;74(4):1043-8; discussion 1048-9.

[44] Edmunds LH, Cohn L.H., eds. Cardiac Surgery in the Adult. Perioperative/ Intraoperative Care. 2004. New York: McGraw-Hill. Pp. 261-548.

[45] Hawe P, Shiell A, Riley T. Complex interventions: how out of control can a randomised controlled trial be? BMJ 2004;328:1561-1563.

[46] 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 Jun;88(6):716-20.

[47] 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.

[48] The Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. NEJM 1991;324:1073-1078.

[49] Vickers AJ, Bianco FJ, Serio AM, Eastham JA, Schrag D, Klein EA, Reuther AM, Kattan MW, Pontes JE, Scardino PT. The surgical learning curve for prostate cancer control after radical prostatectomy. J Natl Cancer Inst 2007; 99: 1171-7.

[50] Neumayer L, Giobbie-Hurder A, Jonasson O, Fitzgibbons R Jr, Dunlop D, Gibbs J, Reda D, Henderson W, for the Veterans Affairs Cooperative Studies Program. Open mesh versus laparoscopic mesh repair of inguinal hernia. N Engl J Med 2004; 350: 1819-27.

[51] Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. Trials 2009, 10:9.

[52] Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. Clin Trials 2004; 1: 421-427.

[53] Cook JA, Ramsay CR, Fayers P. Using the literature to quantify the learning curve. nt J Technol Assess Health Care 2007; 23: 255-260.

[54] Birkmeyer JD, Dimick JB, Birkmeyer NJ. Measuring the quality of surgical care: structure, process, or outcomes? J Am Coll Surg. 2004 Apr;198(4):626-32.

[55] Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. Health Technol Assess. 2001;5(22):1-194

[56] 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: 205–213.

[57] Clavien PA, Barkun J, DeOliveira ML, Vauthey JN, Dindo D, Schulick RD, Santibañes E, Pekolj J, Slankamenac K, Bassi C, Vonlanthen R, Padbury R, Cameron JL, Makuuchi M. The Clavien-Dindo Classification of Surgical Complications: Five-year Experience. (in press).

[58] DeOliveira ML, Winter JM, Schafer M, Cunningham SC, Cameron JL, Yeo CJ, Clavien PA. Assessment of Complications After Pancreatic Surgery: A Novel Grading System Applied to 633 Patients Undergoing Pancreaticoduodenectomy. Ann Surg 2006; 244(6): 931-9

[59] 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 Jul 1;24(19):3178-86. Review.

[60] 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.

[61] 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 Jan 16;6:4.

[62] Meakins JL. Evidence-based practice: new techniques and technology. Can J Surg 2001;44(4):84-5.

[63] Oxman A, Sackett, DL & Guyatt GH. Users' guides to the medical literature. I. How to get started. JAMA 1993 Nov 3; 270 (17) : 2093-2095.

[64] Rothenberger DA. Evidence-based practice requires evidence. Br J Surg 2004; 91: 1387-8.

[65] Rangel SJ, Moss RL. Recent trends in the funding and utilization of NIH career development awards by surgical faculty. Surgery. 2004 Aug;136(2):232-9.

[66] Fischer L, Bruckner T, Diener MK, Kadmon M, Wente MN, Sauerland S, Seiler CM. 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 Jan-Feb;66(1):15-9.

[67] Rahbari NN, Diener MK, Fischer L, Wente MN, Kienle P, Büchler MW, Seiler CM. A concept for trial institutions focussing on randomised controlled trials in surgery. Trials. 2008;9:3.

[68] Maddern G, Boult M, Ahern E, Babidge W. ASERNIP-S: International trend setting. ANZ J Surg. 2008 Oct;78(10):853-8.

[69] 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 Spring;25(2):124-33.

[70] Gordon B, Mackrodt C, Fern E, Truesdale A, Ayers S, Grant A. TheIpswich Childbirth Study: 1. A randomised evaluation of two stage postpartum penned repair leaving the skin unsutured. *Br J Obstet Gynaecol I998;* 105 *435-440.*