Delivering successful randomized controlled trials in surgery:

methods to optimize collaboration and study design

Natalie S Blencowe¹, Jonathan A Cook², Thomas Pinkney³, Chris Rogers⁴, Barnaby C Reeves⁴,
Jane M Blazeby¹,⁵

¹Centre for Surgical Research, School of Social and Community Medicine, University of
Bristol, UK.
²Centre for Statistics in Medicine, Nuffield Department of Orthopedics, Rheumatology and
Musculoskeletal Sciences, University of Oxford, Oxford, UK.
³Academic Department of Surgery, School of Cancer Sciences, Queen Elizabeth Hospital
University of Birmingham, Edgbaston, Birmingham, UK.
⁴Clinical Trials and Evaluation Unit, School of Clinical Sciences, University of Bristol, Bristol,
UK.
⁵Division of Surgery, Head & Neck, University Hospitals Bristol NHS Foundation Trust,
Bristol, UK.

Corresponding author: Professor JM Blazeby  j.m.blazeby@bristol.ac.uk

Word count: 2438

Running head: Methods to optimize randomized controlled trials in surgery
Abstract

Randomized controlled trials in surgery are notoriously difficult to design and conduct, due to numerous methodological and cultural challenges. Over the last five years, several UK-based surgical trial-related initiatives have been funded to address these issues. These include the development of Surgical Trials Centers and Surgical Specialty Leads (individual surgeons responsible for championing RCTs in their specialist fields), both funded by the Royal College of Surgeons of England; networks of research-active surgeons in training; and investment in methodological research relating to surgical RCTs (to address issues such as recruitment, blinding, and the selection and standardization of interventions). This paper discusses these initiatives in more detail and provides exemplar cases to illustrate how the methodological challenges have been tackled. The initiatives have surpassed expectations, resulting in a renaissance in surgical research throughout the UK, such that the number of patients entering surgical RCTs has doubled.

Key words: Surgery, randomized controlled trials, methodology, collaboration, pre-trial work, pilot and feasibility studies
Introduction

In 1996 Richard Horton, editor of the Lancet, published a disparaging editorial about the quality of surgical research, parodying it as a ‘comic opera’.\textsuperscript{1} He highlighted that most published research papers were single surgeon case series, and criticized surgeons’ limited insight about the weakness of this study design in establishing causation. Combined with a lack of collaborative working – ‘...the personal attributes that go to make a successful surgeon differ from those needed for...multicentre research’ - he postulated that specific methodological challenges unique to the evaluation of surgical interventions may explain the lack of randomized controlled trials (RCTs) in this field. Many of these challenges relate to the fact that surgical procedures are complex healthcare interventions,\textsuperscript{2} meaning that ‘unlike 20 milligram tablets, no two surgical procedures are the same’ and achieving standardization of interventions within RCTs is difficult.\textsuperscript{3} Challenges for evaluation arise from surgeons’ personal preferences, as well as their skill levels and ability to learn new techniques. Surgical procedures also involve other team members (i.e. anesthetists, surgical assistants, nurses) and associated pre, peri- and post-operative interventions, all of which may influence the overall outcome of an operation. Moreover, surgery is constantly evolving and it has been said of surgery that it is always ‘too early [for rigorous evaluation such as a trial] until, unfortunately, it’s suddenly too late’.\textsuperscript{4} Other challenges include achieving blinding of surgeons and other caregivers, and appropriate outcome selection; there is a tendency for surgeons to put great emphasis upon early technical adverse events and focus less on longer term patient-centered outcomes.\textsuperscript{5}

In the years following Horton’s editorial, the number and quality of trials in surgery initially remained poor.\textsuperscript{6-8} Less than 10% of published articles in surgical journals are classified as
RCTs and systematic reviews summarizing the quality of RCTs in surgery have shown little improvement.\textsuperscript{6} Whilst the figures for other specialties are not easily available, a conventional evidence base is lacking for much of surgical practice; in the absence of evidence, surgeons have continued to perform operations according to their preferred methods and surgical approaches and strategies for similar conditions vary widely.\textsuperscript{9}

In the US and UK, government funded research bodies (National Institute for Health and National Institute for Health Research) have pledged to improve their nations’ health and wealth through research and increasing access to research studies.\textsuperscript{10} This vision includes the many patients undergoing planned or unscheduled surgical procedures; evidence-based surgical practice is critical and surgical trials should be a key part of achieving this. Over the last five years, the UK has begun a major research investment in this field. This paper describes the main UK based surgical trial-related initiatives that have started and are contributing to this initiative. It also outlines some ways in which the methodological challenges in RCTs in surgery may be tackled.

**Investment in evidence based surgery**

*Creation and support of Surgical Trials Centers*

The Royal College of Surgeons of England has led an initiative aimed at developing an infrastructure for surgical research.\textsuperscript{11} The main components include investment in Surgical Trials Centers and Surgical Specialty Leads (individual surgeons responsible for championing RCTs in their specialist fields). Five centers – selected through a competitive national selection process – were awarded funding in 2013; a sixth joined the network in 2015 and a
seventh is planned for 2016. All are embedded in clinical trials units that are part of a UK network of units meeting specified criteria\textsuperscript{12} and link to the Surgical Specialty Leads (Figure 1). Core funding for the centers facilitates i) collaborations between surgeons and trial methodologists to develop new trials, ii) mentoring of new chief and principal investigators and iii) provision of expertise to engage, educate and encourage teamwork among surgeons across a wide range of hospitals. The centers’ progress is monitored carefully against ambitious targets, which aim (at a minimum) to develop three new investigators and two new studies each year. To date, the initiative has surpassed expectations. In the last three years, 57 RCTs in surgery have been initiated, producing 175 new chief and principal investigators, an increase in the number of hospitals recruiting into surgical studies, and double the number of patients entering surgical RCTs (25,500 in 2014-15 compared with 11,000 in 2011-12).\textsuperscript{13} The creation of Surgical Trials Centers was timely because it coincided with a commissioned call for surgical research led by the National Institute for Health Research. The research teams and centers worked together, resulting in an £18m investment into RCTs in surgery.

\textit{Surgical Specialty leads}

The Royal College of Surgeons provided funding for 15 surgical specialty leads to work with the Trials Centers in developing trials, encouraging networking and providing research-based education for attendings and residents in each surgical specialty. The process was competitive, with each lead appointed in conjunction with leaders from the corresponding specialty society (for example, the upper gastrointestinal lead was selected by the Royal College of Surgeons and leaders of the Association of Upper Gastrointestinal Surgeons). These links have provided opportunities for high profile plenary presentations at national
conferences, as well as a resource for other surgeons to engage and work with. The surgical specialty lead for coloproctology, for example, led a large-scale research prioritization exercise: initially, members of the Association of Coloproctologists of Great Britain and Ireland submitted over 500 research ideas for consideration; these were reduced into a final list of 15 cancer and 10 non-cancer research questions, using modified Delphi methodology involving surgeons, nurses, patients, researchers and industry representatives. Some of the SSLs use their funding to support feasibility work to inform RCTs, and to link with surgical residents to work together in designing and conducting studies.

*Research within surgical training*

In parallel with the renaissance in research led by the Royal College of Surgeons, surgeons in training have reinvented ground-level multicenter research. Surgical residents across the UK have developed collaborative networks, such that nearly all regions of the UK now boast a trainee-led research collaborative. The smaller subspecialties (such as pediatrics, neurosurgery and cardiac surgery) have formed similar networks on a national (rather than regional) scale. Trainee members of collaboratives are working together, both within and across the regions, to design and conduct multicenter randomized and observational studies. Their ability to work together to recruit patients means that large scale studies can increasingly be carried out by mobilizing a large number of people in a short time. The first trainee-led RCT recruited over 700 patients from 21 centers and finished over a year ahead of schedule, setting an ambitious target for future studies which at least one other has achieved. Whether this model is truly transferable to different clinical areas and research studies is uncertain; however, the creation of a workforce who participate and understand trials has huge potential positive benefit for future consultant surgeons and research.
Trainee-led observational studies (which have included as many as 165 centers and over 8,900 patients)\(^{20}\) have quantified differences in practice between surgeons, centers and regions, leading to the identification of evidence gaps and ideas for future RCTs. Another benefit of trainee-led research collaboratives is that they are developing future capacity with expertise in trial design and methodology. This means that even difficult trials can be tackled because of the increased levels of trials expertise and willingness to work in research teams. Thus, the culture of UK surgical research is changing rapidly from being competitive to being collaborative, countering Horton’s views about surgeons’ inability to work together.\(^{21}\)

**Developing and implementing methods to improve the design and delivery of RCTs in surgery**

Surgical trials face particular difficulties. These include challenges with recruitment, intervention design and delivery, and outcome selection and measurement. In the UK, in parallel with the initiatives described above, there has been investment in methodological research to facilitate and enhance RCTs. The UK Medical Research Council created a Network of regional Hubs for Trials Methodology Research.\(^{22}\) One of these - the Collaboration and innovation in Difficult and Complex randomized controlled Trials In Invasive procedures Hub\(^{23}\) - specifically tackles methodological issues in trials of complex interventions, with a focus on surgery. The Hub is co-located with the Royal College of Surgeons Trials Centre in Bristol and is therefore in a unique position to deploy diverse methods to enhance new trials and monitor and evaluate their impact. It also provides a focus of expertise and clinical leadership, bringing surgeons and methodologists together to increase capacity in this field. The Hub works closely with the two registered clinical trials units in Bristol,\(^{24, 25}\) other
regional and national clinical trials units\textsuperscript{26, 27} and the regional trainee research collaborative.\textsuperscript{16} Its major themes in surgical RCTs include the application of methods to optimize recruitment, improve pre-trial pilot and feasibility work, and outcome selection. Benefits include estimating the size of eligible populations and recruitment rates, developing outcome measures, exploring the feasibility of blinding, selecting appropriate interventions and determining their acceptability. These are described below, using examples from completed or ongoing studies.

\textit{Recruitment}

There are many obstacles and hidden challenges to randomizing patients into surgical RCTs.\textsuperscript{28} Historically, recruitment problems were widespread with approximately only 50\% of surgical RCTs reaching scheduled targets.\textsuperscript{29} One reason for this may be that diverse treatments are sometimes compared (e.g. surgical \textit{versus} non-surgical interventions), which creates difficulties due to personal views that one treatment is better than another despite a lack of evidence to confirm this.\textsuperscript{30} It may also occur because surgeons are not familiar with recruiting patients into RCTs. In the Hub, methods have been developed (the Quintet Recruitment Intervention) to overcome these problems.\textsuperscript{31} This uses mixed methods to observe the recruitment process and the way information is communicated to participants. Following this, educational sessions and individual feedback, explaining how information can be communicated more impartially and preferences probed, are provided. This approach has been used successfully in the By-Band-Sleeve study (case study 1), in which there were pre-existing treatment preferences amongst surgeons and patients.

\textit{Blinding of staff and patients}
Blinding of staff and patients in surgical trials is a major challenge. It is, however, key to attempt this wherever possible to minimize performance and ascertainment bias. The use of large dressings to blind patients and staff to the type of incision and surgical access has been recently piloted in feasibility studies\(^{30}\). Whilst this was originally done in early surgical trials\(^{32}\) it has been neglected in many recent studies. In our recent example the success of blinding has also been tested by asking participants and research personnel at the end of a trial whether they were aware of the participant’s allocation.\(^{33}\) The ROMIO (Randomized Oesophagectomy: Minimally Invasive or Open?) study\(^{34}\) (case study 2) demonstrated that blinding of patients and outcome assessors was possible, despite substantial differences between the two procedures under investigation (minimal access surgery versus standard open techniques). Although the most recent CONsolidated Standards Of Reporting Trials statement does not recommend testing for blinding success, it is possible that application of this test in surgical studies is particularly important to educate surgeons as they begin to observe that patients are unaware of their treatment allocation. More work is underway to develop methods to blind outcome assessors in RCTs comparing surgical and non-surgical interventions. A systematic review of RCTs in surgery is being undertaken to identify current methods used to blind staff. Future plans will involve making use of the recently developed collaborative networks (described above) to undertake follow up assessments of patients from each other’s centers.

Selecting and standardizing interventions, and determining their acceptability

Selecting the intervention groups to compare in an RCT requires consideration of existing evidence, current practice and emerging interventions to ensure that the findings of a future large trial would be up to date and relevant.\(^{35, 36}\) Sometimes, however, it is unclear what
constitutes current practice because new techniques are introduced and quickly become widely used – overtaking an evaluation and making results out-of-date before they become available. During the design of a trial, the use of surveys to characterize current practice is recommended in such situations. This approach was used in the Bluebelle study\textsuperscript{37} (case study 3; incidentally another study carried out by trainee surgical collaboratives), collecting data in the operating theatre prospectively across multiple centers for a representative sample of patients. Findings from this multicenter survey were surprising (for example, tissue adhesive was used ‘as a dressing’ in 27% of cases) and when combined with other data sources they resulted in a change in trial group interventions for the main study\textsuperscript{38, 39}

Until recently, little attention has been paid to the complexity of surgical interventions. This complexity is multilayered and has several implications, particularly when determining what constitutes the intervention(s) and comparator(s) of interest and how they are described and standardized in trials. In order for surgeons to adopt trial findings, interventions need to be described in sufficient detail to enable accurate replication;\textsuperscript{40} however, this may not always be necessary or practical to provide.\textsuperscript{41} Careful standardization in a RCT may fail to reflect the inherent variation in clinical practice and consequently reduce the generalizability of the results. A compromise would be to determine an intervention’s ‘active ingredients’ – i.e. those that are thought to optimally influence outcomes or those which are different between the intervention and comparator in a trial – and the degree to which they need to be standardized. The Hub has developed new methods to explore the standardization of surgical interventions during the pilot phase of RCTs. These include development of a typology with which to deconstruct interventions into component parts (in order to consider which may need to be standardized), and qualitative data collection in the operating room (in order to understand how interventions are actually delivered and explore surgeons’
views about standardization). The methods are currently being tested and refined within an RCT and observational study, to explore how the interventions might be described and standardized across different research settings.

It is also important to ascertain the acceptability of the interventions under evaluation in an RCT. This may be particularly relevant in surgery because procedures and technology are constantly evolving and RCTs sometimes compare very different treatment groups, such as invasive versus non-invasive interventions. For example, there is emerging evidence that prophylactic appendectomy may reduce disease relapse in patients with ulcerative colitis. The ACCURE-UK study (case study 4) is conducting preliminary work to establish whether randomization to receive such an invasive intervention (compared with standard pharmacological treatment) is acceptable to patients and clinicians.

**Conclusion**

Surgical trials, particularly multicenter randomized studies, are essential to establish evidence based practice and inform health policy. Their design and conduct has historically been hampered by the lack of a collaborative research culture among surgeons and challenges posed by the surgical setting/context. Over the last five years, UK-based initiatives have made huge strides in establishing an infrastructure to address these methodological and cultural challenges. There has been a subsequent increase in the number of multicenter surgical RCTs, as well as a higher proportion of studies conducting vital preliminary work in order to understand the best ways to overcome barriers to RCTs and improve their quality. Surgical practice in the UK is now firmly on an upward trajectory to be based on evidence, rather than eminence.
Figure 1. Schema of the Royal College of Surgeons of England trials initiative

Surgical Trials Centers

- Oxford Surgical Trials Unit
- Bristol Surgical Trials Centre
- Birmingham Surgical Trials Unit
- North West Surgical Trials Unit
- National Facial & Oral Research Centre
- York Surgical Trials Centre
- Leeds Surgical Trials Centre

Subspecialty leads

- Cardiac
- Upper GI
- Colorectal
- Orthopedics
- ENT
- Breast
- Endocrine
- Plastics
- Head & Neck
- Transplant
- Neurosurgery
- Urology

Regional consultant and trainee collaborative networks
**Case study 1 Assessing and addressing recruitment challenges: the By-Band-Sleeve Study**

**Background**

The prevalence of severe and complex obesity is increasing worldwide and surgery may offer an effective and lasting treatment. Laparoscopic adjustable gastric band and Roux-en-Y gastric bypass surgery are the two main surgical procedures performed, accounting for approximately 80% of all obesity operations in the UK and the United States. However, it is uncertain which operation confers most benefit. An RCT comparing gastric band and bypass was perceived by many to be too difficult to conduct and recruit into because of strong preferences amongst surgeons that influence patient selection for surgery.

**What was done**

A pragmatic RCT with a preceding internal pilot phase was designed. Pilot work aimed to establish whether it was possible to recruit into this surgical trial using integrated qualitative methods. The integrated qualitative study consisted of the following:

(a) A comprehensive process of logging of potential RCT participants through screening and eligibility phases will be used to monitor recruitment. Flow charts will be produced to show the degree of complexity of participation and any variations between centers.

(b) In-depth semi-structured interviews to understand perspectives of participation were conducted with clinical and recruitment staff, and patients. Interviews provided data about: the perspectives of eligible patients; the evidence underlying the RCT, including the importance of the question and the commitment of staff to it, as well as individual clinical equipoise; the application of the protocol in clinical centers and any logistical issues; and suggestions about reasons for recruitment difficulties and potential solutions from those working closely within the RCT. These data formed the basis for feedback to recruiters and to determine the content of the information, and training programs to be initiated in Phase 2 of the RCT.

**What was found**

The qualitative recruitment intervention identified that surgeons needed training to communicate, i) uncertainty, ii) rationale for randomization, iii) the pros and cons of the treatment groups and iv) reasons for participation in the trial. Training was provided with group and individual feedback.

**Implications for the main trial**

The intervention led to increased recruitment in all centers. The progression criteria designed for the internal pilot were met. The funding body therefore approved expansion into the main trial. This has continued and recruitment remains close to being on time and target.

---

**Legend**

RCT = randomized controlled trial
**Case study 2 Achieving blinding of patients: the ROMIO study\textsuperscript{34, 46}**

**Background**

Surgery alone or in combination with chemotherapy or chemoradiation treatment is the mainstay of cure for localized esophageal cancer. There are several surgical approaches, and the past decade has seen a growing interest in minimal access surgical techniques with the potential advantages of causing less tissue trauma and better recovery. The primary outcome for the main trial is currently planned to be patient-reported physical fatigue, measured at several time points during the first three months after surgery. One of the aims of the pilot study was to develop and evaluate feasible, acceptable, and effective methods of blinding patients for the first week postoperatively, so reducing performance and outcome reporting biases.

**What was done**

A pilot study was conducted across two hospitals. Patients undergoing esophagectomy for cancer were randomized to receive totally minimally invasive surgery (laparoscopic abdomen and thoracoscopic chest), totally open surgery or a combination (laparoscopic abdomen and open chest). At the end of the procedure, surgeons placed dressings to cover open and minimally invasive surgery, irrespective of the surgical approach used. Dressings were left in place for five days post-operatively, to enable blinded outcome assessment up to this point. On the sixth day, point patients were asked to guess which operation they had received.

**What was found**

Most patients guessed that they had undergone the ‘combination’ approach irrespective of the group to which they were allocated, suggesting that blinding had been effective. The application (and replacement) of both open and laparoscopic dressings was found to be feasible and in most cases, it was possible to maintain blinding of patients for the first five postoperative days.

**Implications for the main trial**

This method of blinding patients has been successful and will therefore be used to reduce detection bias within the main study.
Case study 3 Selecting the interventions to be evaluated: the Bluebelle study\textsuperscript{37-39}

**Background**

In the UK, wound dressings are routinely applied at the end of abdominal procedures in adults, yet dressings are scarcely used in pediatric populations. A systematic review concluded that there was no evidence to suggest that any dressing significantly reduced the risk of developing a surgical site infection compared with leaving wounds exposed; neither was there any benefit associated with particular dressing types. The National Institute for Health Research therefore commissioned work to examine whether an RCT of dressing vs no dressing would be possible. One of the critical uncertainties was selecting the interventions to be compared, because there are many different dressing types available.

**What was done**

A survey of current practice was undertaken across 20 UK hospitals over a two week period. Surgical trainees collected data in the operating theatre about the types of dressing used after abdominal surgery, and reasons why these were selected. Information relating to operative and patient factors that might influence this choice was also collected.

**What was found**

Data were collected from 724 patients and 1724 wounds. Although most wounds were covered with basic dressings (68%) following wound closure, an unexpectedly high number (27%) were covered with tissue adhesive (the study team were not aware that tissue adhesive was being used routinely). Surprisingly, no differences were found in the dressings used following planned and unscheduled surgery, between different types of surgery (including stoma formation), or according to patient characteristics (such as diabetes, co-morbidities). Surgeons tended to use the dressings that were handed to them at the end of the operation, rather than selecting specific dressings for clinical reasons.

**Implications for the main trial**

Given the survey findings, the main trial will now include tissue adhesive as a separate intervention group, to reflect routine practice. The inclusion criteria will also be widened to include patients undergoing unplanned operations.
### Case study 4 Assessing the acceptability of interventions: the ACCURE-UK Study

**Background**

Ulcerative colitis is a chronic inflammatory condition of the large bowel affecting more than 50,000 people in the UK, of whom around 40% will experience a relapse annually and 25% requiring total colectomy in their lifetime. Ulcerative colitis patients can be treated with maintenance medical therapy; however, the annual disease relapse rate is around 40% which requires escalation to high dose steroid medication with its incumbent risks and toxicity, or progression to a major colectomy operation. Research has shown that the appendix may affect the development and activity of ulcerative colitis. Several small studies in patients with active disease have found that appendectomy reduces relapse, hospitalization and medication usage, with the potential to prevent the need for future major surgery.

The aim of the ACCURE-UK study was to explore whether appendectomy is an acceptable treatment option to ulcerative colitis patients and clinicians and if patients are willing to be randomized to a trial where appendectomy forms one of the treatment groups. It also aimed to establish the safety profile of therapeutic appendectomy in this novel population.

**What was done**

A multicenter randomized external pilot study was undertaken in which adult patients with an established diagnosis of ulcerative colitis, currently in remission, were included and randomised 1:1 to control group (standard medical treatment) or intervention (elective laparoscopic day case appendectomy plus standard medical treatment). Embedded qualitative research interviews, alongside recruitment targets and screening logs, were designed to ascertain acceptability (to both patients and clinicians) and verify the presence of an accessible cohort of eligible patients.

**What was found**

Of 106 patients approached, 60 (56.6%) were willing to be randomized. Overall 53 patients were randomised from 6 study sites. Recruitment took slightly longer than anticipated, but the target cohort of 48 patients was exceeded and the recruitment rate increased over time, with 25 patients recruited in the final two months. Appendectomy was completed as a day case in most cases with minor post-operative complications in four patients.

**Implications for the main trial**

The intervention was shown to be attractive to patients and clinicians, as well carrying minimal surgical morbidity. Patients with ulcerative colitis are willing to be randomized to such a trial and gastroenterologists and surgeons are keen to be involved. We have learnt how to access these patients and how to facilitate and streamline a complicated recruitment pathway. The success of the pilot study suggests that a large phase III trial exploring clinical effectiveness is feasible; a funding application for this full trial is underway.
References

43. Murkin C, Blazeby JM, Blencowe N et al. Response to a pilot single-centre randomized trial: the PATRASTOM trial. Colorectal Dis 2016:n/a-n/a.