

# Selecting and measuring optimal outcomes for randomised controlled trials in surgery

Rhiannon C. Macefield · Caroline E. Boulind ·  
Jane M. Blazeby

Received: 16 October 2013 / Accepted: 25 October 2013 / Published online: 14 November 2013  
© The Author(s) 2013. This article is published with open access at Springerlink.com

## Abstract

**Background** Randomised controlled trials (RCTs) in surgery are complex to design and conduct and face unique challenges compared to trials in other specialties. The appropriate selection, measurement and reporting of outcomes are one aspect that requires attention. Outcomes in surgical RCTs are often ill-defined, inconsistent and at high risk of bias in their assessment and historically, there has been an undue focus on short-term outcomes and adverse events meaning the value of trial results for clinical practice and decision-making is limited.

**Purpose** This review addresses three key problems with surgical trial outcomes—choosing the right outcomes for the trial design and purpose, selecting relevant outcomes to measure from the range of possible outcomes, and measuring outcomes with minimal risk of bias. Each obstacle is discussed in turn, highlighting some suggested solutions and current initiatives working towards improvements in these areas. Some examples of good practice in this field are also discussed.

**Conclusions** Many of the historical problems with surgical trial outcomes may be overcome with an increased understanding of the trial design and purpose and recognition that pragmatic trials require assessments of outcomes that are patient-centred in addition to measurement of short-term outcomes. The use of core outcome sets developed for specific surgical interventions and the application of novel methods to blind outcome assessors will also improve outcome measurement and reporting. It is recommended that surgeons work together with trial methodologists to integrate these approaches into RCTs in surgery. This will facilitate the

appropriate evaluation of surgical interventions with informative outcomes so that results from trials can be useful for clinical practice.

**Keywords** Outcomes · Randomised controlled trial · Surgery · Trial design · Blinding

## Introduction

Randomised controlled trials (RCTs) in surgery are uncommon and often more complex and difficult to design and implement than trials of pharmaceutical interventions. The reasons for this include challenges with recruitment (due to lack of clinical equipoise or treatment preference), difficulties with standardisation and delivery of interventions between trial centres or individual surgeons (particularly if the interventions under evaluation are new) [1, 2], and challenges concerning the appropriate selection, measurement and reporting of outcomes.

An outcome (or endpoint) is a direct or indirect measurement of the effect of an intervention on a participant's clinical or functional status [3, 4]. The purpose of outcome measurements in an RCT is to provide information about the effect of an intervention under evaluation compared to a standard procedure or control. In a trial, participants are randomised to receive different interventions with the aim of negating the effect of confounding factors. In this way, any observed differences in outcome between the treatment groups can be attributed to the effect of the intervention and not to a baseline characteristic of the participant or other potentially influential variable. When designing a clinical trial, it is necessary to define a trial hypothesis of the expected effect of the intervention on a primary outcome (the outcome of greatest interest). This allows the calculation of a sample size which is essential to inform the number of centres in the trial and the length of trial recruitment. In some trials, it may be advantageous to have a composite primary outcome comprised of more than

---

R. C. Macefield · C. E. Boulind · J. M. Blazeby  
Bristol Centre for Surgical Research, School of Social and  
Community Medicine, University of Bristol, Canynge Hall,  
39 Whatley Road, Bristol BS8 2PS, UK

J. M. Blazeby (✉)  
Division of Surgery, Head and Neck, University Hospitals Bristol  
NHS Foundation Trust, Bristol, UK  
e-mail: J.M.Blazeby@bristol.ac.uk

one outcome of interest. This can be advantageous when event rates are low, to allow for smaller sample size calculation and increased statistical efficacy, or when the choice between several important outcomes is arbitrary [5]. Limitations of using a composite outcome, however, include the need for a more complex statistical analysis plan and ensuring that all component outcomes have equal importance to patients [6, 7]. Secondary outcomes are measured to evaluate the additional effects of the intervention and answer other relevant questions [8]. They may also be chosen for exploratory purposes in order to develop a hypothesis for future research. Outcomes may be classified with regard to their type (primary, secondary, or exploratory), and whether they are assessed objectively or influenced by patient or clinician judgement [9]. They may also be categorised by the domain of measurement. Clinical outcomes may be defined as outcomes measured in clinical practice such as survival/mortality or surgical complications and adverse events (e.g. anastomotic leak or deep vein thrombosis). Clinical outcomes can be completely objective (e.g. death) or they can include a professional assessment or observer interpretation (e.g. wound infection). Other types of outcomes may be patient-reported outcomes (PROs); assessments made directly by the patient themselves rather than an observer, such as symptom severity or reports of functional health status. Most PROs are recorded in self-completed questionnaires with validated rating scales that are standardised and scored and are often referred to as measures of health-related quality of life (HRQL). Whilst some outcomes may be referred to as HRQL outcomes, this term is confusing, and it is recommended that outcomes are described as PROs (measured by patients themselves), or observer assessed outcomes (measured by clinicians/researchers), rather than the ill-defined term HRQL—which may be measured by a patient or an observer. Surrogate outcomes are outcomes that predict another clinically meaningful endpoint of interest [10] and may be used because they are more practical or less invasive, for example, the measurement of coagulation factors or albumin as a surrogate marker of liver function in patients following liver surgery [11]. Other types of outcomes include hospital or ‘process’-related factors such as length of hospital stay or frequency of tests, and assessments of cost and resource use (Box 1).

Many of these different types of outcomes can be assessed once or on several occasions following surgery. The timing of outcome assessment varies in trials, and can be classified as during or after surgery, as short, medium or long-term. Short-term assessments are often made during the hospital stay and longer-term assessments maybe years after the intervention. In the selection of outcomes for RCTs in surgery, it is necessary to consider these different types of outcomes, who will assess them, and the appropriate time to assess them in light of the research question and trial design.

Historically, surgeons have selected and assessed outcomes that focus on short-term clinical events and those related to the surgical process such as complication rates and length of hospital stay [12]. This may be because these outcomes reflect aspects of patient outcome and recovery that are of importance to the surgeons designing the study, and reflect the quality of the delivery of the intervention and the surgeons’ technical skill [13]. They are also events that require early detection and treatment by surgeons, which may result in serious problems for patients if untreated. Whilst it is important to assess and report these short-term clinical outcomes, many surgical studies do not examine the longer-term benefits and harms of surgical interventions or consider outcomes from a patient perspective [14]. In addition, undue focus on operative complications may be unhelpful as these outcomes are often rare, and if trials are not powered sufficiently to exclude differences in uncommon complications, the data can be misleading. There is, therefore, a need for better standards of outcome selection, measurement and reporting in RCTs in surgery. This paper will consider some of the key challenges and suggest solutions to these problems, highlighting current initiatives working towards improvements in these areas.

### Challenges for outcome measurement in RCTs in surgery

The methodological and practical challenges for the selection and measurement of outcomes in surgical trials include choosing the right outcome for the trial design and purpose, selecting relevant outcomes to measure from the range of possible outcomes and measuring them with minimal risk of bias.

#### Choosing relevant outcomes to suit the trial design and purpose

In order to select the most appropriate outcomes for any trial, it is first necessary to have a good understanding of the research hypothesis. This will determine the trial design and the appropriate outcomes to be measured. Trials can be broadly classified as explanatory or pragmatic in design. Explanatory (efficacy) trials are designed to test whether an intervention works under ideal conditions (performed by an experienced surgeon, for example, in a single centre with standardised pre- and post-operative care) and for these reasons, the intervention is delivered within a very restrictive protocol often on a highly select group of patients [2, 15]. Outcomes in explanatory trials are selected to provide information on whether the intervention works within these restricted conditions and tend to focus

on measurable clinical or biological symptoms or markers [16]. For surgical trials, these are typically short-term clinical or process measures that supply data about the safety of the intervention and its immediate health benefits and risks. Examples include measured changes in intra-operative variables (e.g. length of operation, blood loss) or short-term adverse events and in-hospital morbidity (e.g. need for pain killers, length of hospital stay) (Table 1, [17]). Pragmatic clinical (effectiveness) trials, on the other hand, are designed to assess whether an intervention works in routine clinical practice and everyday settings. The intervention is delivered under more flexible conditions and by practitioners with a range of expertise to reflect typical practice. Members of the research team, such as the surgeons, often have more scope to make choices about the exact procedure performed, or the approach used (depending on the research intervention). The sample population is usually less restricted and in combination, these factors aim to make the results generalisable to a wide range of clinical environments. Pragmatic trials often provide information on the longer-term health gains and harms of an intervention and results are used to inform patients, clinicians and policy makers when making decisions on treatment. A wide spectrum of outcomes can be assessed during a pragmatic trial and an assessment of resource use may also be included. The majority of RCTs in surgery aim to determine if the intervention should be performed in standard clinical practice and therefore, a pragmatic trial design is often most appropriate [18]. However, reviews of the surgical literature show that the number of high-quality pragmatic trials is few and often focus on short-term and clinical measures providing uncertain value for patients, surgeons and other decision makers [19–23]. Although there is a paucity of studies, there are examples of well designed and conducted pragmatic RCTs in surgery with appropriate outcome measures (Table 1). There are also examples of trials using composite outcomes combining short-term measures and including some longer-term assessments [24]. However, more high-quality RCTs in surgery are required with better trial design and outcome measures to allow the results to be relevant to routine clinical practice. This will require that surgeons involved in trial design to consider which outcomes are important to patients and include assessment of these in trials as well as the standard measures used to evaluate surgery.

Multicentre, pragmatic trials require surgeons to work together and to involve patients and other health-care professionals in the trial design process at an early stage. The National Institute for Health Research (NIHR) in the UK funds an advisory group to support greater public and patient involvement in research [25] and the James Lind Alliance charitable organisation aims to bridge the gap between patients and researchers [26]. A similar initiative in Canada, the Patient-

Centred Outcomes Research Institute, encourages the involvement of patients to shape the research agenda [27]. In the UK, recent funding opportunities from the NIHR and Royal College of Surgeons of England for Surgical Trial Centres [28] are also promoting collaborative working, and there are trainee research collaboratives whose purpose is to bring together surgical trainees and offer the opportunity to run multicentre trials [29, 30]. Drives to bring together surgeons and methodologists and to educate the surgical community in clinical trial design have been generated. Examples of this include the international IDEAL collaboration [31], the UK Medical Research Council Hubs for Trials Methodology Research [32] and the American College of Surgeons Continuous Quality Improvement Surgical Research Committee [33]. In Germany, the Study Centre for the German Surgical Society has similar aims to increase the number of well-designed multicentre surgical trials [34]. These initiatives should lead to improvements in the design of randomised trials of surgical interventions.

#### Selecting and reporting appropriate outcome measures for RCTs in surgery

After establishing the trial design and relevant outcomes, the next challenge for RCTs in surgery is to clearly define and select the outcomes to be measured from the range of possibilities. In surgical studies, the approach to outcome assessment has historically been assorted, inconsistent and ill-defined. A systematic review of adverse events after gastrointestinal surgery, for example, identified 56 different definitions and measures for anastomotic leak [35] and up to 10 different measures for mortality have been used in studies after oesophagectomy [23]. In colorectal cancer surgery, a systematic review found 766 different clinical outcomes were assessed, with inconsistency in selection, measurement and reporting [36]. This variation across trials and lack of standardisation is problematic for systematic reviews and meta-analyses of surgical interventions as outcomes are often not comparable, meaning synthesis and amalgamation of findings is seldom possible. In addition to inconsistent outcome measurements, other problems for RCTs in surgery are caused by the multiplicity of outcomes that are measured in a single trial. As a result, authors of trial reports do not always include the full range of outcomes that have been measured, often focusing on outcomes with ‘interesting’ or statistically significant results [37, 38]. Such selective reporting of outcomes leads to outcome reporting bias and causes further problems for systematic reviews by distorting the available evidence [39–41]. Outcome reporting bias is a particular issue for PROs, where numerous multiple health domains are measured within a single questionnaire and those of particular interest are often not specified a priori or when PROs are secondary outcomes.

**Table 1** Short- and long-term outcomes commonly used in examples of RCTs in surgery

Trial	Aim	Primary outcome	Secondary short-term outcomes		Secondary long-term outcomes			Comments
			Clinical	PRO	Clinical	PRO	Process Economic	
CaCSH <i>Santarius et al. 2009</i> [17]	To investigate post-operative drainage in the management of chronic subdural haematoma after burr-hole evacuation	Rate of reoperation to treat recurrent chronic subdural haematoma	30-Day mortality Hospital stay (days) MRS and GCS at discharge Gross focal neurological deficit at discharge	N/A	Mortality at 0.5 years MRS and GCS at 0.5 years Gross focal neurological deficit at 0.5 years Mobility status at 0.5 years	N/A N/A N/A N/A	N/A N/A N/A N/A	Single-centre RCT with outcomes focusing on clinical endpoints. This limits generalisability
CLASICC <i>Jayne et al. 2007</i> [61]	To evaluate the technical and oncological safety and efficacy of laparoscopically assisted surgery for colorectal cancer	Several were defined: Resection margins Proportion of Dukes' C2 tumours In-hospital mortality 3-year overall survival, disease-free survival and local recurrence rates	30-Day and 3-month complication rates Transfusion requirements	HRQL at 2 weeks and 3 months	3- and 5-year distant recurrence rates 3- and 5-year wound/port site recurrence rates 5-year overall survival 5-year disease-free survival 5-year local recurrence	N/A N/A N/A N/A	N/A N/A N/A N/A	This very pragmatic trial had a sample size selected to provide an overall picture of several short- and long-term outcomes. It was not powered for specific endpoints
CORONIS <i>Abalos et al. 2013</i> [24]	To assess whether five surgical caesarean techniques were associated with improved outcomes for women and babies	A composite outcome: Death Maternal infectious morbidity* Further operative procedures Blood transfusion up to 6-week follow-up visit	Individual components of composite outcome Pain Interventions for post-partum haemorrhage Still birth Other severe maternal morbidity Apgar score <3 at 5 min Laceration of baby during caesarean section Death of baby at 6 weeks Operation length Hospital stay Admission to ITU Hospital readmission within 6 weeks	HRQL at 1 and 3 months	Assessed in follow-up study [62]: Reproductive status, subsequent pregnancies, death or serious morbidity of child at 3 years Women's health and mortality (requested at face to face interview) at 3 years	N/A N/A N/A N/A	N/A N/A N/A N/A	Multicentre, multinational pragmatic RCT design. Authors state outcomes were specifically selected to provide guidance for clinical decision-making and composite components were chosen based on primary objective of the trial
EVAR <i>Greenhalgh et al. 2010</i> [63]	To compare the long-term effects of endovascular and open repair of large aneurysms	All-cause mortality Aneurysm-related mortality within 30 days	Graft-related complications Graft-related interventions 30-day operative mortality	HRQL at 1 and 3 months	Graft-related complications Graft-related interventions Adverse events Renal function (up to 8 years)	HRQL at 1 year	Cost-effectiveness based on QALY (EQ-5D) Resource use annually up to 8 years	Multicentre pragmatic trial designed to provide long-term data, although HRQL outcomes not measured for duration of trial
<i>King et al. 2006</i> [64]	To compare short-term outcomes of laparoscopic and open resection of colorectal cancer within an enhanced recovery programme	Hospital stay	Convalescent hospital stay Readmission hospital stay Major morbidity Requirement for opioid analgesia Antiemetic administration Performance indicators for mobility/strength at 2 and 12 days, 6 and 12 weeks	HRQL at 2 and 6 weeks Resource use at 2 weeks and 3 months	N/A	N/A	N/A	Single-centre RCT with small sample size designed to assess short-term outcomes only

**Table 1** (continued)

Trial	Aim	Primary outcome		Secondary short-term outcomes		Secondary long-term outcomes			Comments
		Clinical	PRO	Clinical	PRO	Clinical	PRO	Economic	
REFLUX Grant et al. 2013 [65]	To evaluate laparoscopic fundoplication surgery with medical management for the treatment of (GORD)	Validated PRO with the REFLUX questionnaire	Sleep and oxygen saturation at 2, 6 12weeks SF36, EQ-5day & REFLUX HRQL score at 3 months Surgical complications	Mortality Use of anti reflux drugs at 3 months and 1 year	SF36 and EQ-5D score at 1, 2, 3, 4 and 5 year REFLUX HRQL and REFLUX symptom score at 2, 3, 4 and 5 years	N/A	N/A	N/A	Multicentre pragmatic RCT with PRO primary outcome. N.B. REFLUX questionnaire has several scales to form an overall score
SYNTAX Mohr et al. 2013 [66]	Comparison of CABG with PCI for the treatment of patients with left main coronary disease three-vessel disease	Composite: rate of MACCE at 1 year (all-cause mortality, stroke, myocardial infarction, repeatrevascularisation)	Rate of MACCE at 1 month	Rate of MACCE at 0.5, 1, 3 and 5-year post procedure	HRQL at 1-month post procedure	N/A	Cost and cost-effectiveness at 1, 3 and 5 years	N/A	Multicentre, multinational pragmatic RCT. Practical interpretation limited as components of composite outcome were not weighted to reflect their importance or relative frequency

N/A not assessed; CABG coronary artery bypass graft surgery; EQ-5D EuroQol questionnaire; GCS Glasgow coma scale; GORD gastro-oesophageal reflux disease; HRQL health-related quality of life; MACCE major adverse cardiac and cerebrovascular events; MRS modified Rankin score; PCI percutaneous coronary intervention; PRO patient-reported outcome; QALY quality-adjusted life year; SF-36 short-form 36 Health survey

\*Defined as one or more of: antibiotic use for maternal febrile morbidity during postnatal hospital stay, antibiotic use for endometritis, wound infection, or peritonitis up to the 6-week follow-up visit

† Defined as haemorrhage requiring blood transfusion, reoperation, readmission, anastomotic leak, wound dehiscence and sepsis requiring high-dependency support

**Table 2** Methods to blind patients, surgeons and outcome assessors in examples of RCTs in surgery

Trial	Interventions	Primary outcome	Type of outcome	Personnel blinded	Methods used for blinding
Raviele et al. 2004 [52]	Permanent pacing versus placebo for recurrent tilt-induced vasovagal syncope	Recurrence rate of syncope at 1 year assessed by patient diaries	PRO	Patients Surgeons Clinical staff	Pacemakers implanted into both treatment arms using the same method; a single person in each centre was responsible for the programming of the pacemaker to ON in active mode with rate drop response, or OFF in the inactive mode.
Sung et al. 2003 [53]	Endoscopic treatment versus irrigation alone for non-bleeding vessels or adherent clot in gastroduodenal ulceration	Recurrence of bleeding before discharge and at 30 days	Independent adjudication panel	Patients Clinical staff Outcome assessors	Intervention group underwent routine endoscopic treatment for bleeding ulcer including irrigation, suction, heater probe, or mini-snare. The control group received irrigation of ulcer base but no manipulation with heater probe, snare or suction; post-operative care by blinded clinical team; endoscopist not involved in post-operative care or outcome assessment; criteria for re-bleed pre-defined and assessed by blinded panel.
Moseley et al. 2002 [49]	Arthroscopic lavage and debridement versus arthroscopic lavage only versus sham procedure for osteoarthritis of the knee	Knee pain at 24 months assessed using Knee-Specific Pain Scale (developed for this trial)	PRO	Patients Nurses Clinical staff	Patients allocated to lavage and debridement or lavage only underwent general anaesthesia and endotracheal intubation. Patients in the control arm underwent short-acting intravenous sedation and opioid analgesia with spontaneously breathed oxygen-enriched air. Patients allocated to lavage and debridement received a standard arthroscopic procedure. Patients allocated to lavage only underwent an identical procedure except that no debridement was performed (unless an unstable meniscal tear was identified in which case this was excised). For patients in the control arm, the knee was prepped and draped as usual and the 1-cm incisions performed. No instrumentation was performed but the surgeon manipulated the knee as per arthroscopic debridement. Post-operative care was provided by blinded clinical staff according to a standardised pathway; the surgeon was not involved in post-operative care or outcome assessment.
Vitek et al. 2003 [54]	Pallidotomy versus medical therapy for Parkinson's disease	Average change in unified Parkinson's disease rating scale at 6 months	Assessor reported	Independent assessors	Two independent outcomes assessors blind to treatment allocation collected all outcome data; all patients wore hats to mask scars during baseline and follow-up outcomes collection; there was no contact between assessors and participants between follow-up appointments; assessors were not involved in the routine care of participants; patients were asked not to inform assessors of treatment allocation.
Quinn et al. 2002 [55]	Suturing versus conservative management of hand lacerations	Cosmetic appearance at 3 months	Assessor reported	Independent assessors	Assessment of wounds made photographically with no knowledge or contact with participant at 3 months after treatment by two independent doctors.
Gervaz et al. 2010 [56]	Laparoscopic versus open (lower midline) approach for sigmoid colectomy to treat diverticular disease	Composite pain assessed daily for 4 days; time to first flatus/bowel movement	PRO	Patients Nursing staff	Sterile, opaque dressing to cover entire lower abdomen for 4 days
Ezra et al. 2004 [57]	Vitrectomy alone versus vitrectomy plus autologous serum transfer versus conservative management for full thickness macular hole	Composite closure of macular hole and visual acuity (time point unclear)	Assessor reported	Independent assessors	Closure of hole accessed via digitalised photographs of funduscopy and digitalised images of fluorescein angiography. These assessments were made separately as vitrectomy is obvious on funduscopy. This way partial blinding between the arms was maintained because serum transfer cannot be seen; blinded assessment by a separate assessor of visual acuity

PRO patient-reported outcome

Core outcome sets (COSs) are one solution to the problems caused by multiple outcome selection and outcome reporting bias. Core outcome sets are a collection of the important (core) outcomes to be measured and reported in all pragmatic trials of a specific disease or condition, and their use and application allows for the results from multiple trials to be readily combined and compared. Core outcomes are agreed by consensus between key stakeholders such as patients, health-care professionals, and members of funding bodies. The Core Outcome Measures for Effectiveness Trials (COMET) Initiative in the UK supports the development and application of COSs and holds a database of all existing work in this area [42]. Several COSs to use in trials of surgical intervention are being developed, for example in oesophageal, colorectal and head and neck cancer as well as other diseases outside oncology [42–44].

#### Assessing and measuring outcomes with minimal risk of bias

Once the outcomes to be measured are agreed upon and defined, it is necessary to ensure that they are assessed appropriately to reduce the risk of bias. Bias can exaggerate treatment effect, thereby making trial results unreliable and problematic when implementing for clinical practice [9]. Consideration must be given to how outcomes are measured and by whom. It is also important that the tools used to assess these outcomes are developed, tested, and validated properly, as the use of inappropriate tools result in unreliable or invalid data. This applies to any outcome measure being used in a trial, whether it be a clinical measure of anastomotic leak, for example, or a PRO.

There are several forms of bias from which surgical trials are particularly at risk, including performance bias and ascertainment bias. Performance bias results from differences in the way patients in the two treatment arms are managed post-operatively. It can have an effect on all types of outcomes, and is a particular problem in trials of surgical interventions because of the intimate role played in both delivery of the intervention and post-operative care provision by the operating surgeon. Ascertainment bias is a term used to describe collectively observer bias (when outcomes are reported by an outcome assessor) and reporting bias (for PROs), and can be a particular problem in trials whose primary outcome is a PRO. Both performance and observer bias can occur on a conscious or sub-conscious level and might result from pre-existing expectations about the efficacy of the treatments under evaluation, a lack of clinical equipoise or the influence of knowledge about treatment allocation. One method that can be used to minimise the risk of both performance and ascertainment bias is blinding (or masking). Blinding is used in RCTs to hide treatment allocation from various members of the research team or trial

participants. In the ideal setting, blinding is maintained until all primary outcome data are collected. The decision about who should be blinded in a particular trial depends on the trial design, interventions and available funding. In pharmacological trials, it is common to use a matched placebo to blind staff and patients. In the case of surgical trials, however, using a placebo or sham surgery is often difficult or impossible because it may be considered unethical to subject patients to general anaesthesia and a mock operation for solely research purposes, although such examples exist [45–51]. The complex nature of surgery can also make blinding of staff difficult because of the necessity for input from a large multidisciplinary team and it is frequently impossible for surgeons to be blinded because of their role in delivery of the intervention. In this case, where ever possible, it is recommended that the surgeon does not collect the outcome data and that it is collected by someone blinded to the intervention type.

Despite these challenges for surgical RCTs, novel methods may be effectively employed to blind some of the research team. Examples where this has been successfully achieved are illustrated in Table 2 [49, 52–57]. In a trial whose primary outcome measure is a PRO, it is important to consider whether it is possible to blind the patient from the treatment allocation, although this may also be impossible. It may be more practical in this situation to blind outcome assessors who are ideally independent from the central research team. For example, an adjudicating committee may be used to review and assess digitalized clinical, photographic or imaging data with no details of the patient or intervention available to them [57].

#### Summary and conclusion

Well designed and conducted pragmatic RCTs are essential for the evaluation of surgical interventions. Central to this is the need to select, define, measure and report outcomes that are relevant to the trial design, valid and reliable, and to assess these outcomes with minimal ascertainment and performance bias. This can be achieved with surgeons and methodologists working closely together. The need to improve and increase the number of surgical trials has been internationally recognised and a cultural shift in the approach to trial design and participation is taking place. The development and application of core outcome sets for pragmatic RCTs in surgery will reduce many of the problems associated with the inconsistency and multiplicity of outcomes that have historically been assessed in RCTs in surgery and this is being addressed by work of the COMET initiative. Finally, the challenge to reduce and eliminate bias in surgical trials is being tackled by innovative methods to perform blinding to ensure that outcomes

are fairly assessed and measured and effects of the intervention are not over or under estimated. These combined approaches and efforts will ensure surgical interventions are appropriately evaluated with informative outcomes, so that results can be useful for clinical practice.

### Box 1. Types of outcomes and outcome measures used in RCTs in surgery

#### Clinical outcome measure

An outcome measured in clinical practice, e.g. survival or surgical complications such as anastomotic leak, including the specific measurement variable (e.g. systolic blood pressure), analysis metric (e.g. change from baseline, final value, time to event), method of aggregation (e.g. median, proportion) and timepoint [58].

#### Patient reported outcome (PRO)

An outcome reported directly by the patient themselves without interpretation from an observer [59]. Examples include assessments of health status and quality of life (e.g. physical ability, symptom severity). PROs are typically recorded in a self-completed questionnaire.

#### Patient-reported outcome measure

A measure of a PRO including domain (e.g. anxiety), specific measurement tool (e.g. name of questionnaire) and other levels of specification required for clinical outcomes (analysis metric and method of aggregation) [60].

#### Hospital/process-related outcome

A metric related to the organisation or individual involved in the patients' care rather than the effect of the intervention on the patient's health, e.g. length of hospital stay or number of tests conducted.

#### Resource use measure

A metric to quantify the cost of care, including direct financial expenses as well as staff time.

#### Surrogate outcome

A measure that is not of direct practical importance but is believed to reflect an outcome that is important; often a physiological or biochemical marker that can be relatively quickly and easily measured, and that is taken as being predictive of an important clinical outcome [3].

#### Composite outcome

An outcome that consists of two or more component outcomes [5].

#### Primary outcome

The outcome of greatest importance [3].

#### Secondary outcome

An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes [3].

#### Explanatory outcome

Outcomes that are measured to provide additional information about an intervention, but may be without an a priori hypothesis.

**Conflicts of interest** None.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

1. Developing and evaluating complex interventions: new guidance. (2008). [www.mrc.ac.uk/complexinterventionsguidance](http://www.mrc.ac.uk/complexinterventionsguidance). Accessed 18 September 2013
2. Cook JA (2009) The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials* 10:9
3. Glossary of Cochrane terms. <http://www.cochrane.org/glossary>. Accessed 18 September 2013
4. Wang D, Bakhai A (2006) *Clinical trials: a practical guide to design, analysis, and reporting*. Remedica, London
5. Cordoba G, Schwartz L, Woloshin S, Bae H, Gotzsche PC (2010) Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *British Medical Journal* 341. doi:10.1136/bmj.c3920
6. Ross S (2007) Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol* 196(2):e1–e6
7. Montori VM, Permyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, Alonso J, Akl EA, Domingo-Salvany A, Mills E, Wu P, Schunemann HJ, Jaeschke R, Guyatt GH (2005) Validity of composite end points in clinical trials. *Br Med J* 330(7491):594–596. doi:10.1136/bmj.330.7491.594
8. CONSORT Glossary. <http://www.consort-statement.org/resources/glossary/>. Accessed 18 September 2013
9. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JPA, Schulz KF, Beynon R, Welton N, Wood L, Moher D, Deeks JJ, Sterne JAC (2012) Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 16(35):1–82. doi:10.3310/hta16350
10. Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JAC, Taylor RS (2013) Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *British Medical Journal* 346. doi:10.1136/bmj.f457
11. Mpabanzi L, van Mierlo KMC, Malago M, Dejong CHC, Lytras D, Damink S (2013) Surrogate endpoints in liver surgery related trials: a systematic review of the literature. *HPB* 15(5):327–336. doi:10.1111/j.1477-2574.2012.00590.x
12. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, Seiler CM, Balliol C (2009) Surgical innovation and evaluation 2. Challenges in evaluating surgical innovation. *Lancet* 374(9695):1097–1104
13. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Nicholl J, Balliol C (2009) Surgical innovation and evaluation 3. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 374(9695):1105–1112
14. Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A (2013) IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *Br Med J* 346:f2820. doi:10.1136/bmj.f2820
15. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ, Chalkidou K (2009) A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 62(5):464–475. doi:10.1016/j.jclinepi.2008.12.011
16. Patsopoulos NA (2011) A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 13(2):217–224
17. Santarius T, Kirkpatrick PJ, Ganesan D, Chia HL, Jalloh I, Smielewski P, Richards HK, Marcus H, Parker RA, Price SJ, Kirolos RW, Pickard JD, Hutchinson PJ (2009) Use of drains versus no drains after burr-hole evacuation of chronic subdural haematoma: a randomised controlled trial. *Lancet* 374(9695):1067–1073. doi:10.1016/s0140-6736(09)61115-6

18. Lilford R, Braunholtz D, Harris H, Gill T (2004) Trials in surgery. *Br J Surg* 91(1):6–16. doi:10.1002/bjs.4418
19. Pibouleau L, Boutron I, Reeves BC, Nizard R, Ravaud P (2009) Applicability and generalisability of published results of randomised controlled trials and non-randomised studies evaluating four orthopaedic procedures: methodological systematic review. *British Medical Journal* 339. doi:10.1136/bmj.b4538
20. Ellis C, Hall JL, Khalil A, Hall JC (2005) Evolution of methodological standards in surgical trials. *ANZ J Surg* 75(10):874–877. doi:10.1111/j.1445-2197.2005.03554.x
21. Jacquier I, Boutron I, Moher D, Roy C, Ravaud P (2006) The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement—a systematic review. *Ann Surg* 244(5):677–683. doi:10.1097/01.sla.0000242707.44007.80
22. Potter S, Brigg A, Whiting PF, Cawthorn SJ, Avery KNL, Donovan JL, Blazeby JM (2011) Reporting clinical outcomes of breast reconstruction: a systematic review. *J Natl Cancer Inst* 103(1):31–46. doi:10.1093/jnci/djq438
23. Blencowe NS, Strong S, McNair AGK, Brookes ST, Crosby T, Griffin SM, Blazeby JM (2012) Reporting of short-term clinical outcomes after esophagectomy: a systematic review. *Ann Surg* 255(4):658–666. doi:10.1097/SLA.0b013e3182480a6a
24. Abalos E, Addo V, Brocklehurst P, El Sheikh M, Farrell B, Gray S, Hardy P, Juszczak E, Mathews JE, Masood SN, Oyarzun E, Oyieke J, Sharma JB, Spark P (2013) Caesarean section surgical techniques (CORONIS): a fractional, factorial, unmasked, randomised controlled trial. *Lancet* 382(9888):234–248. doi:10.1016/s0140-6736(13)60441-9
25. INVOLVE Supporting public involvement in NHS, public health and social care research. <http://www.invo.org.uk/>. Accessed 30 September 2013
26. The James Lind Alliance. <http://www.lindalliance.org/>. Accessed 18 September 2013
27. Fleurence R, Selby JV, Odom-Walker K, Hunt G, Meltzer D, Slutsky JR, Yancy C (2013) How the patient-centered outcomes research institute is engaging patients and others in shaping its research agenda. *Health Aff* 32(2):393–400. doi:10.1377/hlthaff.2012.1176
28. Royal College of Surgeons of England. <http://www.rcseng.ac.uk/>. Accessed 18 September 2013
29. Surgical trainee collaboratives. <http://www.asit.org/resources/collaboratives>. Accessed 20 September 2013
30. Bhangu A, Koliass AG, Pinkney T, Hall NJ, Fitzgerald JE (2013) Surgical research collaboratives in the UK. *Lancet* 382:1091–1092
31. Surgical research: the reality and the IDEAL (2009) *Lancet* 374(9695):1037–1037
32. Medical Research Council Network of hubs for trials methodology research. <http://www.methodologyhubs.mrc.ac.uk/>. Accessed 18 September 2013
33. American College of Surgeons Continuous Quality Improvement Surgical Research Committee. <http://www.facs.org/cqi/src/>. Accessed 18 September 2013
34. Rahbari NN, Diener MK, Fischer L, Wente MN, Kienle P, Buechler MW, Seiler CM (2008) A concept for trial institutions focussing on randomised controlled trials in surgery. *Trials* 9. doi:10.1186/1745-6215-9-3
35. Bruce J, Krukowski ZH, Al-Khairi G, Russell EM, Park KGM (2001) Systematic review of the definition and measurement of anastomotic leak after gastrointestinal surgery. *Br J Surg* 88(9):1157–1168. doi:10.1046/j.0007-1323.2001.01829.x
36. Whistance RN, Forsythe RO, McNair AG, Brookes ST, Avery KN, Pullyblank AM et al (2013) A systematic review of outcome reporting in colorectal cancer surgery. *Colorectal Dis* 15(10):e548–e560
37. Chan AW, Altman DG (2005) Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *Br Med J* 330(7494):753. doi:10.1136/bmj.38356.424606.8F
38. Hannink G, Gooszen HG, Rovers MM (2013) Comparison of registered and published primary outcomes in randomized clinical trials of surgical interventions. *Ann Surg* 257(5):818–823. doi:10.1097/SLA.0b013e3182864fa3
39. Dwan K, Altman DG, Amaiz JA, Bloom J, Chan AW, Cronin E et al (2008) Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 3(8)
40. Dwan K, Kirkham JJ, Williamson PR, Gamble C (2013) Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. *BMJ Open* 3(6)
41. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR (2010) The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *British Medical Journal* 340. doi:10.1136/bmj.c365
42. COMET (Core Outcome Measures in Effectiveness Trials) Initiative. <http://www.comet-initiative.org/>. Accessed 19 September 2013
43. Macefield RC, Jacobs M, Blencowe NS, Korfage IJ, Nicklin J, Brookes ST, Sprangers MAG, Blazeby JM (2011) The case for a HRQL core outcome set: outcome reporting bias in oesophageal cancer studies. *Trials* 12(Suppl 1):A77
44. Whistance RN, Blencowe NS, Blazeby JM (2012) The need for standardised outcome reporting in colorectal surgery. *Gut* 61(3):472–472. doi:10.1136/gutjnl-2011-300676
45. Schulz KF, Grimes DA (2002) Blinding in randomised trials: hiding who got what. *Lancet* 359(9307):696–700. doi:10.1016/s0140-6736(02)07816-9
46. Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P (2007) Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 4(2):e61. doi:10.1371/journal.pmed.0040061
47. Boutron I, Tubach F, Giraudeau B, Ravaud P (2003) Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *J Am Med Assoc* 290(8):1062–1070. doi:10.1001/jama.290.8.1062
48. Boutron I, Tubach F, Giraudeau B, Ravaud P (2004) Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 57(6):543–550. doi:10.1016/j.jclinepi.2003.12.010
49. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP (2002) A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 347(2):81–88. doi:10.1056/NEJMoa013259
50. Freed CR, Greene PE, Breeze RE, Tsai WY, DuMouchel W, Kao R, Dillon S, Winfield H, Culver S, Trojanowski JQ, Eidelberg D, Fahn S (2001) Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N Engl J Med* 344(10):710–719. doi:10.1056/nejm200103083441002
51. Wolsko PM, Eisenberg DM, Simon LS, Davis RB, Walleczek J, Mayo-Smith M, Kaptchuk TJ, Phillips RS (2004) Double-blind placebo-controlled trial of static magnets for the treatment of osteoarthritis of the knee: results of a pilot study. *Altern Ther Health Med* 10(2):36–43
52. Raviele A, Giada F, Menozzi C, Specca G, Orazi S, Gasparini G, Sutton R, Brignole M (2004) A randomized, double-blind, placebo-controlled study of permanent cardiac pacing for the treatment of recurrent tilt-induced vasovagal syncope. The vasovagal syncope and pacing trial (SYNPACE). *Eur Heart J* 25(19):1741–1748. doi:10.1016/j.ehj.2004.06.031
53. Sung JJ, Chan FK, Lau JY, Yung MY, Leung WK, Wu JC, Ng EK, Chung SC (2003) The effect of endoscopic therapy in patients receiving omeprazole for bleeding ulcers with nonbleeding visible vessels or adherent clots: a randomized comparison. *Ann Intern Med* 139(4):237–243
54. Vitek JL, Bakay RA, Freeman A, Evatt M, Green J, McDonald W, Haber M, Barnhart H, Wahlay N, Triche S, Mewes K, Chockkan V, Zhang JY, DeLong MR (2003) Randomized trial of pallidotomy versus medical therapy for Parkinson's disease. *Ann Neurol* 53(5):558–569. doi:10.1002/ana.10517

55. Quinn J, Cummings S, Callahan M, Sellers K (2002) Suturing versus conservative management of lacerations of the hand: randomised controlled trial. *Br Med J* 325(7359):299
56. Gervaz P, Inan I, Perneger T, Schiffer E, Morel P (2010) A prospective, randomized, single-blind comparison of laparoscopic versus open sigmoid colectomy for diverticulitis. *Ann Surg* 252(1):3–8. doi:10.1097/SLA.0b013e3181dbb5a5
57. Ezra E, Gregor ZJ (2004) Surgery for idiopathic full-thickness macular hole: two-year results of a randomized clinical trial comparing natural history, vitrectomy, and vitrectomy plus autologous serum: Morfields Macular Hole Study Group Report no. 1. *Arch Ophthalmol* 122(2):224–236. doi:10.1001/archophth.122.2.224
58. Chan A-W, Tetzlaff JM, Altman DG, Dickersin K, Moher D (2013) SPIRIT 2013: new guidance for content of clinical trial protocols. *Lancet* 381(9861):91–92. doi:10.1016/s0140-6736(12)62160-6
59. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD (2013) Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *J Am Med Assoc* 309(8):814–822. doi:10.1001/jama.2013.879
60. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC (2011) The ClinicalTrials.gov results database—update and key issues. *N Engl J Med* 364(9):852–860. doi:10.1056/NEJMsa1012065
61. Jayne DG, Guillou PJ, Thorpe H, Quirke P, Copeland J, Smith AM, Heath RM, Brown JM (2007) Randomized trial of laparoscopic-assisted resection of colorectal carcinoma: 3-year results of the UK MRC CLASICC Trial Group. *J Clin Oncol* 25(21):3061–3068. doi:10.1200/jco.2006.09.7758
62. CORONIS: International study of caesarean section surgical techniques: the follow-up study. (2012). <https://www.npeu.ox.ac.uk/files/downloads/coronis-follow-up/CORONIS-Follow-up-Protocol-V5-Nov-2012.pdf>. Accessed 04 October 2013
63. Greenhalgh RM, Brown LC, Powell JT, Thompson SG, Epstein D, Sculpher MJ (2010) Endovascular versus open repair of abdominal aortic aneurysm. *N Engl J Med* 362(20):1863–1871. doi:10.1056/NEJMoa0909305
64. King PM, Blazeby JM, Ewings P, Franks PJ, Longman RJ, Kendrick AH, Kipling RM, Kennedy RH (2006) Randomized clinical trial comparing laparoscopic and open surgery for colorectal cancer with an enhanced recovery programme. *Br J Surg* 93(3):300–308. doi:10.1002/bjs.5216
65. Grant AM, Cotton SC, Boachie C, Ramsay CR, Krukowski ZH, Heading RC, Campbell MK (2013) Minimal access surgery compared with medical management for gastro-oesophageal reflux disease: five year follow-up of a randomised controlled trial (REFLUX). *Br Med J* 346:f1908. doi:10.1136/bmj.f1908
66. Mohr FW, Morice MC, Kappetein AP, Feldman TE, Stahle E, Colombo A, Mack MJ, Holmes DR Jr, Morel MA, Van Dyck N, Houle VM, Dawkins KD, Serruys PW (2013) Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical SYNTAX trial. *Lancet* 381(9867):629–638. doi:10.1016/s0140-6736(13)60141-5