European Surgical Research

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606
Published online: September 18, 2014

0014–312X/14/0534–0095$39.50/0
www.karger.com/esr

Invited Review

Clinical Research in Surgery: Threats and Opportunities

Wim P. Ceelen

Department of Surgery, Ghent University Hospital, Ghent, Belgium

Key Words
Research · Randomized trial · Surgery · Evidence

Abstract
Surgery is a discipline which profoundly affects human integrity. Therefore, there is an ethical and scientific imperative that surgical practice depends on the best possible trial-based evidence. Traditionally, the quality and quantity of clinical research have been lagging behind other disciplines in clinical medicine. However, recent collaborative initiatives, such as the IDEAL framework which tests surgical innovation, international registries, and quality assurance platforms, the development of modified randomized controlled trials and alternative trial designs as well as the impending reforms of the regulatory framework surrounding non-pharmaceutical interventions and devices offer significant and timely opportunities to enhance the relevance of clinical research in surgery. Here, we provide an overview of the current state of clinical research in surgery, identify possible obstacles, and discuss realistic and emerging solutions that have the potential to change the way surgical research is organized, funded, and translated to the patient's benefit.

Introduction

The surgical community may be credited with some of the most innovative breakthroughs in medicine, having since a major impact on the life and health of patients. Examples of pioneering efforts include trauma care, solid tumor oncology, angiogenesis research, transplantation, and many more. In most of these areas, however, the exponential increase in the

Wim P. Ceelen, MD, PhD
Department of Surgery 2K12 IC, UZ Gent
De Pintelaan 185
BE–9000 Ghent (Belgium)
E-Mail wim.ceelen@ugent.be

KARGER

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

depth and complexity of basic research has led to the current situation that the surgical research community is increasingly underrepresented in this field, even when it concerns traditional 'surgical' topics [1].

In clinical research, however, the landscape is less barren. Since the inflammatory hyperbole by the Editor-in-Chief of *The Lancet*, who chose to ridicule surgical research, surgeons have risen to the challenge and are increasingly well organized and productive in clinical outcomes research [2]. However, several obstacles remain and hurdles need to be taken in order to unleash the full research potential of a discipline which is unique in its immediate life-saving potential, its access to human anatomy, physiology, and tissue, and its familiarity with advanced technology.

Here, an overview is provided of the current state of clinical research in surgery, the obstacles and threats impeding further progress, and possible measures and policy choices that have the potential to elevate the quality and quantity of surgical clinical research to the level they deserve.

The Current Clinical Research Landscape in Surgery

Randomized Clinical Trials: Why Bother?

Since the landmark first randomized blinded clinical trial published in the *British Medical Journal* in 1948 evaluating the use of streptomycin for pulmonary tuberculosis, the randomized controlled trial (RCT) has been established as the best possible design to evaluate medical interventions [3]. Its preeminence stems from the fact that only a well-executed RCT can, with a sufficient degree of certainty, establish whether any observed treatment effect is due to the prescribed intervention rather than due to known or unknown confounders. Starting in the 1990s, the efforts of several multidisciplinary panels aiming to establish the minimal reporting standards of RCTs in the biomedical literature ultimately resulted in the publication of the CONSORT (Consolidated Standards of Reporting Trials) guidance in 1996 [4]. The guideline was subsequently updated in 2010 [5]. Numerous reports have shown a consistent improvement in the reporting quality of RCTs since the introduction of the CONSORT guidelines [6]. Nevertheless, progress has been less optimal in certain subdomains such as the reporting of adverse events [7]. Also, reporting the quality of nonrandomized trials, such as phase II oncology studies, remains poor, even in journals with strict editorial policies [8]. In order to address the specific circumstances and requirements of non-RCTs that evaluate nonpharmacological interventions including surgery, the CONSORT group recently proposed an extension (nonpharmacological treatment, NPT) aiming to improve the reporting in this domain [9].

Current Quality of Clinical Research in Surgery

Recently, Adie et al. [10] have scrutinized 150 RCTs of surgical interventions published mainly in surgical journals. They found that less than half of these trials described essential methodological details such as sample size calculation (45%), random sequence generation (43%), allocation concealment (45%), or blinding (37%). In solid organ transplantation, the RCTs addressed on average only 47% of the CONSORT items [11]. Adherence to the NPT extension of the CONSORT guideline seems even poorer: Nagendran et al. [12] found that, among 54 surgical trials, 8 CONSORT items were identified with less than 30% overall compliance, and 7 of these were specific to the CONSORT-NPT extension. Gray et al. [13] observed that, although the adherence of surgical trials to the CONSORT guideline significantly improved between 2004 and 2010, items specific to the CONSORT-NPT extension remain underreported.

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

More generally, surgical intervention trials often fail to reach the required scientific standards for clinical trials. Wenner et al. [14] conducted a systematic review of publications from 1999 to 2008 evaluating invasive therapeutic procedures. They found numerous shortcomings such as the mentioning of funding sources (36% failure), complete statistical power calculations (45% failure), and factors required to assess the generalizability of trial results including the number and inclusion criteria of interventionists. In addition, trial discontinuation has recently been demonstrated to be far more common in surgical compared to medical RCTs (43 vs. 27%, p = 0.001), and this was significantly associated with nonpublication [15]. Also, a poor choice of the control arm, with properties that do not compare with general practice, may unduly dilute or inflate any effects detected [16].

Why the Quality of Clinical Research in Surgery Needs Improving

First and foremost, there is an ethical duty not to submit subjects to invasive treatments that are of unproven benefit and potentially harmful. Unfortunately, there have been surgical procedures that were widely adopted before it became apparent that they cause more harm than benefit. As an example, extracranial-intracranial vascular bypass to reduce the risk of ischemic stroke was widely performed based on a single case report. In 1985, however, a randomized trial demonstrated that extracranial-intracranial bypass actually increased the risk of fatal and nonfatal stroke compared to medical therapy alone, and the procedure has been abandoned ever since [17]. Similarly, internal mammary artery ligation was a popular procedure to treat angina pectoris until a randomized comparison with sham surgery reported no benefit in postoperative angina and performance metrics [18]. Other notable examples include the Halsted radical mastectomy, kidney decapsulation to treat hypertension, and uterine suspension. A recent comparison of randomized and nonrandomized studies in breast cancer surgery found that, depending on the metric used, in 20–40% of outcomes the effect estimates differed more than twice between both study types, calling into question the validity of evidence derived from nonrandomized comparisons [19]. Similar findings were reported by Peinemann et al. [20], who performed a systematic review of methodological studies examining the effect of the study type on the reported results. They found that effect sizes between RCTs and non-RCTs were statistically different in 35% of the studies. Clearly, the abundance of retrospective chart reviews and noncontrolled studies therefore negatively impacts the evidence base for surgical practice.

Why Has Adoption of the RCT Been Slow in Surgical Research?

Methodological Issues

Ethical Obstacles – Lack of Equipoise

Subjecting a therapeutic intervention to a prospective comparative trial presupposes that a genuine uncertainty exists about the expected outcome. There has to exist, both from the point of view of the subject and from that of the physician, a sense of 'equipoise', which implicates that both treatment arms are perceived as potentially beneficial. When comparing invasive interventions, or invasive treatments with medical therapy alone, safeguarding the principle of equipoise quickly becomes problematic. Surgeons may be reluctant to include patients in a comparative trial because they believe the benefits of an untested procedure are self-evident. A recent survey of beliefs and attitudes amongst practicing surgeons showed that they often did not feel in equipoise based on a limited appreciation of the methodological weakness of nonrandomized studies, little understanding of pragmatic trial design, and limited belief in the value of RCTs for generating high-quality data to change clinical practice [21]. This will particularly apply to procedures that are already widely disseminated in

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

clinical practice before being subjected to a RCT; examples include sentinel node biopsy in breast cancer and laparoscopic surgery for cancer of the colon. Similarly, patients are increasingly well informed and are unlikely to consent if chance gets to decide whether or not they will receive the procedure for which they have sought the specific expertise of the treating surgeon. In fact, patient refusal is recognized as one of the main barriers to recruitment in surgical oncology trials [22]. Patient preference is often reinforced by an active promotion of novel (but unproven) interventions for reasons of commercial recruitment or personal professional prestige. For similar reasons, trials comparing surgical with medical interventions suffer from high rates of dropout and crossover. In a trial comparing surgical versus nonoperative treatment of lumbar disc herniation, only 50% of the patients assigned to surgery actually received surgery, while 30% of those assigned to the nonoperative treatment received surgery [23]. Because of this important crossover rate, the authors were unable to conclude anything about the superiority or equivalence of the treatments based on an intention-to-treat analysis.

Standardization and Learning Curve Effects

Unlike pharmaceutical treatment, the efficacy of a surgical intervention depends on the skills, preferences, and experience of the operator, which adversely affects the possibility to standardize the intervention. Therefore, when comparing two or more surgical interventions in a RCT, any measured effect will be confounded by the variability in the surgical skills between the participants. Not only is there a potential variation in proficiency between surgeons, but also, and even more importantly, the skills of an individual surgeon tend to improve with time and experience. This learning curve effect, when insufficiently addressed in the planning and execution of a trial, may invalidate the trial findings, specifically in studies evaluating technically complex or high-risk procedures. As an example, a recent analysis of the learning curve for robotic radical prostatectomy showed that, for a high-volume surgeon changing from open to robotic surgery, the functional and oncological (surgical margins) outcomes were actually inferior to those of open prostatectomy during the first 100–150 cases [24]. The reality of learning curves in surgery does not only pose methodological but also ethical problems and controversies [25]. In addition, imbalances in surgeon expertise may confound the outcome of RCTs comparing different types of surgery. Gastric cancer trials comparing standard with extensive (D2) resection found an increased complication rate in the extended surgery group, suggesting that the true performance of D2 gastrectomy may be obscured by differences in surgeons' experience [26]. Conversely, the results of a trial investigating a procedure that was only performed by expert surgeons may lack external validity, i.e., they might not be mirrored by the larger surgical community. Similarly, the choice of appropriate and standardized outcome measures is often haphazard in surgical trials. A systematic review of adverse events after gastrointestinal surgery, for example, identified 56 different definitions and measures for anastomotic leakage and up to 10 different measures for mortality after esophagectomy [27, 28]. Selective outcome reporting is therefore a particularly worrying source of bias in surgical trials.

Lack of Blinding

Successful blinding, which can be applied to investigators, participants, and outcome assessors, is critical in preventing the introduction of bias and may also aid in the compliance and retention of participants. When a surgical intervention is studied, it is obvious that the investigators cannot be blinded. However, the outcome assessors may be blinded; it has been shown that the analysis of surgical trial results by unblinded assessors introduces significant bias [29]. Blinding of participants to a trial comparing surgical techniques may be logistically difficult to implement, although it certainly is possible. A classic example is the trial by Majeed

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

et al. [30] comparing small-incision open with laparoscopic cholecystectomy, during which patients were blinded by the application of a similar, large wound dressing in both groups. Although theoretically ideal, the use of sham (placebo) surgery as a comparator is limited to the rare cases where the potential risk of the sham intervention is outweighed by the potential harm of the active surgery. A recently published RCT has assigned patients with symptoms of degenerative meniscal tear to arthroscopic partial meniscectomy or sham surgery [31]. Interestingly, the outcome after partial meniscectomy was no better than that after a sham surgical procedure.

Lack of Funding

Similar to what has happened in the basic science disciplines, much of the clinical research that is directly or indirectly related to the treatment of surgical disease is being performed and published by nonsurgeons. As an example, clinical trials of neoadjuvant and/or adjuvant therapy in digestive cancer are rarely undertaken under the leadership of a surgical investigator. In parallel, the public funding of surgical grant applications is becoming increasingly difficult. In the USA, the National Institutes of Health (NIH) funding rates of academic surgeons have been declining relative to their nonsurgical colleagues over the past decades [32]. Similarly, the success rate of NIH funding for research proposed by cardiothoracic surgeons has steeply declined [33]. One of the primary causes seems, however, to be the low number of applications submitted by surgeons. Here, the current economic climate and the managed-care model have resulted in the present situation that many academic surgical departments are unable to afford sufficient protected research time to their staff.

Also, in contrast to pharmaceutical research, there is little or no interest from the commercial industry to fund surgical randomized trials. It has been repeatedly suggested that national funding bodies should acknowledge this structural disadvantage and cater for surgery-driven research [34].

Lack of Training

The successful completion of surgical trials is hampered by the fact that very few surgical investigators received any formal training in research methods or clinical epidemiology. From a recent survey among surgical trialists in the USA, it became apparent that the professional development as a trialist was a difficult, haphazard, and inefficient process [35]. Also, given the traditional 'master-pupil' model that continues to form the foundation of surgical training, most surgeons developing an expertise do so according to a 'school' or method that they tend to internalize and prefer over time, which may cause reluctance to change to an unfamiliar approach for the necessity of a clinical trial.

Regulatory Aspects

The realization that the current legal framework regulating the use of medical interventions and/or devices is inadequate was hastened by several notorious public health disasters. In the USA, the FDA requires that moderate-risk (class II) devices are cleared by the demonstration of 'substantial equivalence' to a previously cleared 'predicate' device. The underlying assumption is that if a new device is equivalent to a previous similar device, it will be at least as safe and effective as that device, and no clinical studies are required [36]. In the European Union, medical device regulation is the responsibility of each of the 27 member states. Class II devices are reviewed by one of the 76 'notified bodies' accredited by each country's competent authority. If this (for-profit) firm, paid in part by the manufacturer, deems the device satisfactory, it grants the Conformité Européenne mark, and the manufacturer can then market the device throughout the European Union. As a consequence, the European Union regulatory system has recently been portrayed as 'fragmented, privatized, and largely

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

opaque; safety is dealt with in an unsatisfactory way and efficacy not at all' [37]. Obviously, this regulatory lack of the need for any clinical safety or efficacy data for most surgical devices is one of the root causes of the poor evidence available.

Approaches to Enhance the Quality of Clinical Research in Surgery

Methodological Approaches

Alternative Trial Designs

Although the RCT is considered to provide the highest grade of scientific evidence, alternative designs may be acceptable in those surgical trials where the methodological rigors required may be beyond reach for practical, logistical, or ethical reasons. In fact, it has been suggested that the results of prospective well-designed cohort or case-control studies yield similar estimates of the magnitude of the treatment effects as RCTs [38]. In rare cases of life-saving novel procedures, the need for a comparative trial may even be dropped altogether. However, the risk of bias increases as the magnitude of the treatment effect becomes smaller; it has been suggested that at least a 5- to 10-fold improvement in outcome is needed for a RCT to be unnecessary [39]. Some surgical clinical research questions will require an alternative design because a RCT is impractical or would require excessive funding or time to complete. As a minimum, such alternative 'quasi-experimental' designs should mimic the methodological assets of a RCT as much as possible, including the use of standard definitions, the clear delineation of inclusion criteria, methods, an analytical plan, and the description of the quality assurance methods used. For skill-dependent interventions in general, it has been suggested to abandon the 'pyramid' hierarchical model of evidence quality in favor of a circular model (fig. 1) which addresses the conflict between internal validity (scientific and methodological rigor relating mainly to the avoidance of bias) and generalizability (relevance for the wider surgical community) [40, 41].

Prospective Parallel-Group Nonrandomized Trial. In this design, a cohort of patients undergoing a novel procedure is prospectively followed and compared to another group which receives standard treatment. Apart from the randomization and blinding, all other methodological requirements of a RCT should be fulfilled. Careful attention should be given to the elimination of bias due to baseline imbalances between both groups in terms of clinical and pathological variables such as age, comorbidity, and previous treatments. In addition to a multivariate regression analysis, propensity score matching is increasingly used in order to achieve an adequate matching of baseline characteristics [42].

Interrupted Time Series. This design aims to monitor a treatment outcome parameter of interest over time before and after the introduction of a novel technique or device. Obviously, this design is susceptible to changing indications over time and to learning curve effects; both should therefore be addressed in the study design.

Stepped Wedge Design. In a stepped wedge design, an intervention is administered sequentially to the participants (or groups of participants) over a number of time periods (fig. 2). The order in which the participants receive the intervention is determined at random and, by the end of the study, all individuals or groups will have received the experimental intervention. This design is particularly relevant when it is predicted that the intervention will do more good than harm, or when a real or perceived lack of equipoise may preclude adequate recruitment [43].

Evaluating Surgical Innovation: The IDEAL Framework

The IDEAL framework has recently been developed as a methodological concept for the systematic evaluation of the different time points (idea, development, exploration, assessment,

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

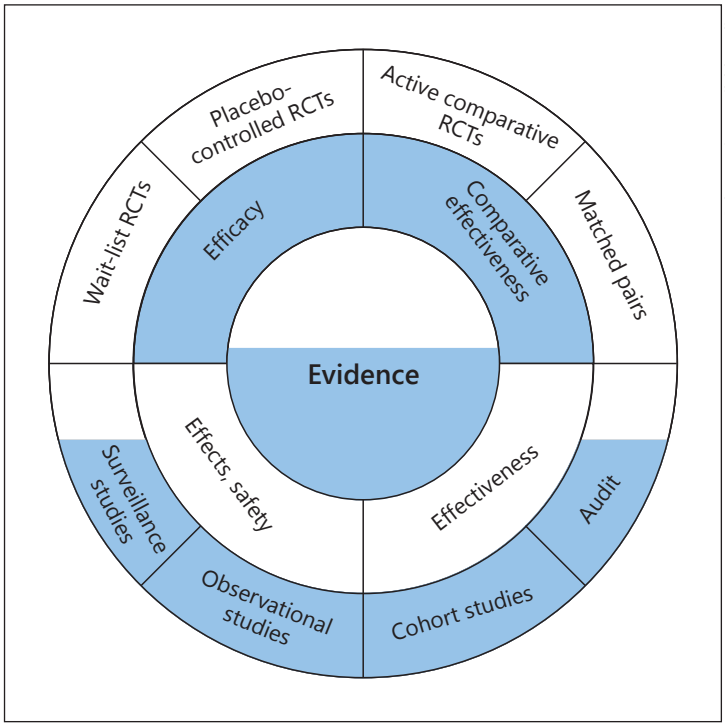

Fig. 1. Circular model of experimental methods. Trials that test specifically for efficacy (upper half of the circle) have to be complemented by observational nonexperimental methods (lower half of the circle) that are more descriptive in nature and describe real-life effects and applicability. The latter can range from retrospective audit studies and prospective case series to one-armed and multiple-armed cohort studies. Matched-pairs studies can be conducted as experimental studies by forming first pairs and then randomizing them or as quasi-experimental studies by forming pairs from naturally occurring cohorts according to matching criteria. Shading indicates the complementarity of experimental and quasi-experimental methods of internal and external validity. Reprinted with permission from Walach et al. [41].

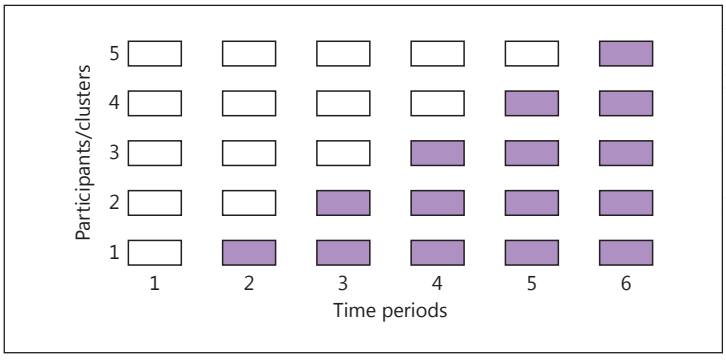

Fig. 2. Example of a stepped wedge study design. Shaded cells represent intervention periods and blank cells control periods. Each cell represents one data collection point. Reprinted with permission from Brown and Lilford [43].

and long-term evaluation) during the introduction of innovative invasive techniques and/or devices [44–48]. At every step, the demands of scientific rigor are reconciled to a maximal extent with the specific circumstances and challenges associated with evaluating skill- and device-dependent interventions (table 1).

Whenever possible, novel techniques or devices should be evaluated in silico (modelling, simulation, and virtual reality) and in vivo (animal models); this step may be regarded as

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

Table 1. Stages in the evaluation of surgical innovation according to the IDEAL approach (modified from [45])

	(1) Idea	(2a) Development	(2b) Exploration	(3) Assessment	(4) Long term
Purpose	proof of concept	development	learning	assessment	surveillance
Patients	single digit; highly selected	few; selected	many; may expand to mixed; broadening indication	many; expanded indications (well-defined)	all eligible
Surgeons	very few; innovators	few; innovators and some early adopters	many; innovators, early adopters, early majority	many; early majority	all eligible
Output	description	description	measurement; comparison	comparison; complete information for non-RCT participants	description, audit, quality assurance
Intervention	evolving; procedure inception	evolving; procedure development	evolving; procedure refinement; community learning	stable	stable
Method	structured case reports	prospective development studies	research database; explanatory or feasibility RCT	RCT with or without modifications; alternative designs	registry; database; rare case reports
Outcomes	proof of concept	safety; technical success	safety; short-term outcomes	specific longer-term outcomes; cost-effectiveness	rare events; quality control
Ethical approval	sometimes	yes	yes	yes	yes

Stage 0. In Stage 1 (idea or innovation), the procedure is tested for the first time in humans by a single surgeon or a small group of surgeons. There should be a clear clinical need driving the innovation, and informed consent is mandatory although formal approval by the local ethical committee may not always be required. Based on the results of this proof-of-concept study, the intervention may be further developed (Stage 2a) in a small group of patients. In contrast to currently prevailing practice, a prospective development protocol should be established, including detailed inclusion criteria, details of the intervention, and outcome(s) to be measured. All prospective studies should be registered with online repositories such as the US NIH clinicaltrials.gov database. Once the safety of the intervention has been established and its technical execution has been adequately standardized, further exploration (Stage 2b) is warranted, aiming in the first place to establish preliminary clinical efficacy data that may serve as the basis for a later RCT. Since, at this stage, multicenter participation is preferred, monitoring and addressing any learning curve effects will be of importance. When this phase of development suggests an important clinical benefit, the intervention should be subjected to a formal RCT comparing it to the current standard of care (Stage 3). In some circumstances, a RCT may not be ethical, practical, or feasible and modified or alternative designs may be considered. Even when the clinical effectiveness of a novel intervention has been established by high-quality evidence, the results should be monitored on a long-term basis (Stage 4), usually in the form of a (web-based) registry. The aim of such surveillance is twofold: first, comparable to pharmaceutical postmarketing follow-up, it allows to detect rare and unanticipated harms. Second, it provides an important tool to monitor the quality of the delivered intervention, including the possibility to provide feedback to the individual participants.

Implementation of Standardized Outcome Measures

The inherent bias associated with using nonstandardized outcome measures and selective outcome reporting may be addressed using standard definitions, such as the Dindo-Clavien classification of postoperative complications, and by the use of core outcome sets (COS). COS are collections of important outcomes to be measured and reported in all pragmatic trials of a specific disease or condition. They are ideally agreed upon by consensus between key stake-

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

holders such as patients, healthcare professionals, and funding bodies. The Core Outcome Measures for Effectiveness Trials (COMET) initiative, launched in 2010, brings together researchers interested in the development and application of agreed standardized sets of outcomes [49]. Several COS applicable to surgical trials are being developed, for example in esophageal, colorectal, and head and neck cancer [50]. In addition, more emphasis is being put on patient-reported outcomes, which may more accurately reflect the effectiveness and quality of surgical interventions compared to the traditional endpoints of morbidity and mortality [51, 52].

Addressing and Controlling the Learning Curve

Confounding of the results of a RCT by the surgeon's learning curve may be avoided at the design level and at the analysis level. At the first level, participants may be required to demonstrate a minimal proficiency in the technique to be tested. In the COLOR II trial comparing open with laparoscopy-assisted rectal cancer surgery, participants were required to submit unedited recordings of 5 consecutive laparoscopic procedures for assessment or were observed by an expert [53]. Similar volume and/or quality criteria can be required from the participating centers as well as from the individual surgeons. A second approach is to use sophisticated statistical techniques, such as Bayesian hierarchical models, in order to adjust the trial results for any learning curve effects [54]. As a minimum, reports of surgical trials should specify the prior expertise of the participating surgeons.

Collaboration as a Recipe for Success

Collaborative efforts between institutions are much more likely to generate high-quality research [55]. In addition, internationally coordinated efforts may allow surgeons experienced in research and academic evaluation to provide for better and more surgical care worldwide, including in the developing countries. In the Netherlands, a long-standing tradition of nationally coordinated research efforts has resulted in the successful completion of a steady output of very-high-quality practice-changing multicenter trials. In Germany, the Study Centre of the German Surgical Society (SDGC) was created in order to facilitate planning, conducting, and analyzing randomized multicenter trials. As of 2012, 5 RCTs have been completed and another 7 are ongoing, recruiting a total of 2,500 patients in over 100 trial centers [56]. In the UK, The Royal College of Surgeons and several partners have established a network of surgical trial units located in Bristol, Oxford, London, Birmingham, Liverpool, and Manchester [57]. Similarly, drives to bring surgeons and methodologists together and to educate the surgical community in clinical trial design have been generated. Examples include the international IDEAL collaboration, the UK Medical Research Council Hubs for Trials Methodology Research, and the American College of Surgeons Continuous Quality Improvement Surgical Research Committee. In the USA, a major initiative involves the Alliance for Clinical Trials in Oncology, formed in 2011 after the merger of the American College of Surgeons Oncology Group, the Cancer and Leukemia Group B, and the North Central Cancer Treatment Group. This initiative will allow to expand surgical research from efficacy trials (the traditional subject of collaborative cancer trial groups) to comparative effectiveness research including registries, selective retrospective studies, and pragmatic RCTs [58]. Another recent initiative from the US is the establishment of the Department of Veterans Affairs Cooperative Studies Program Network of Dedicated Enrollment Sites (NODES), which is hoped to foster interest in collaborative surgical trials [59]. In addition, international collaborations offer unique opportunities to enhance the relevance of surgical research through disease- or procedure-specific registries, such as the LiverMetSurvey database, and international quality enhancement initiatives including the Worldwide Esophageal Cancer Collaboration and the European Registration of Cancer Care (EURECCA) program [60–62].

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

Recruitment, Training, and Motivation of Academic Surgeons

Thanks to long overdue and welcome developments such as working time directives, the increasing representation of women, and changing work-life balance priorities, young surgeons no longer answer to the stereotypic prototype surgeon who passionately spends his time from dawn till (long after) dusk in the operating room and any rare spare time on the golf course. Academic surgical departments should endeavor to recruit the best and brightest medical students, and offering the possibility to participate in research is an important asset. Although only a minority of surgeons in training will eventually pursue an academic career, a minimal education and training in clinical research methods and epidemiology should be provided to all, and more advanced training to some. Within the academic departments, the right balance should be struck between clinical and research-oriented staff, but a 'critical mass' of research interest and activity is an absolute requirement to maintain a culture of scientific inquiry.

Multidisciplinary and Interdisciplinary Collaboration

In order to face the challenges posed by the lack of funding and manpower, there is now an increasing trend towards interdisciplinary collaboration, where clinical and basic researchers share a common research theme. Examples include the numerous institutes for biomedical engineering worldwide, where the collaboration of engineers, surgeons, and basic scientists enhances the depth and quality of surgical research. Also, the integration of multiple academic disciplines and nonacademic individuals into clinical research may help to tackle real-world patient care issues, create more practice-based evidence, and help close the gap between trial-derived knowledge (best practice) and actual patient care [63].

Editorial Policies

Journal editors have been instrumental in the development of minimal standards in trial reporting such as the CONSORT and STROBE guidelines [64]. Similar demands should be required by surgical journals when confronted with research evaluating surgical innovation. As a minimum, details should be provided regarding the prospective nature of the protocol, inclusion criteria, participant experience, and blinded outcome assessment. In parallel, high-quality evidence of novel surgical devices or interventions should be required by other stakeholders such as funding agencies, patient advocacy groups, regulatory bodies, and professional societies.

Reform of the Regulatory Framework

In September 2012, the European Commission proposed an important reform of the regulation of medical devices which consists, amongst other elements, of much stricter requirements for clinical evidence in order to ensure patient and consumer safety [65]. The draft of the novel Medical Devices Regulation, which is expected to have an important impact on the clinical evaluation of surgical innovation, was voted by the European Parliament in April 2014. Similarly, in the USA, the FDA has announced several initiatives in order to strengthen device regulation, including the development of new guidance and the enhancement of staff training [66].

Conclusions

Clinical research in surgery is at a crossroads: either it is destined for the fate that begot much of the basic laboratory research, or it embraces the many opportunities that have the potential to change surgery into a truly evidence-based discipline. Increasingly and rightly so,

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

patients as well as funding bodies demand that surgical practice, which is usually life-changing, sometimes life-saving, but can also be a potential threat when inappropriately administered, is based on sound and best possible scientific evidence. It is up to us to take up this challenge.

Disclosure Statement

The author does not have any conflicts of interest to disclose.

References

1 Smythe WR: The future of academic surgery. Acad Med 2010;85:768–774.
2 Horton R: Surgical research or comic opera: questions, but few answers. Lancet 1996;347:984–985.
3 STREPTOMYCIN treatment of pulmonary tuberculosis. Br Med J 1948;2:769–782.
4 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF: Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–639.
5 Schulz KF, Altman DG, Moher D; CONSORT Group: CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18.
6 To MJ, Jones J, Emara M, Jadad AR: Are reports of randomized controlled trials improving over time? A systematic review of 284 articles published in high-impact general and specialized medical journals. PLoS One 2013;8:e84779.
7 Sivendran S, Latif A, McBride RB, Stensland KD, Wisnivesky J, Haines L, Oh WK, Galsky MD: Adverse event reporting in cancer clinical trial publications. J Clin Oncol 2014;32:83–89.
8 Grellety T, Petit-Moneger A, Diallo A, Mathoulin-Pelissier S, Italiano A: Quality of reporting of phase II trials: a focus on highly ranked oncology journals. Ann Oncol 2014;25:536–541.
9 Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P; CONSORT Group: Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008;148:295–309.
10 Adie S, Harris IA, Naylor JM, Mittal R: CONSORT compliance in surgical randomized trials: are we there yet? A systematic review. Ann Surg 2013;258:872–878.
11 Liu LQ, Morris PJ, Pengel LH: Compliance to the CONSORT statement of randomized controlled trials in solid organ transplantation: a 3-year overview. Transpl Int 2013;26:300–306.
12 Nagendran M, Harding D, Teo W, Camm C, Maruthappu M, McCulloch P, Hopewell S: Poor adherence of randomised trials in surgery to CONSORT guidelines for non-pharmacological treatments (NPT): a cross-sectional study. BMJ Open 2013;3:e003898.
13 Gray R, Sullivan M, Altman DG, Gordon-Weeks AN: Adherence of trials of operative intervention to the CONSORT statement extension for non-pharmacological treatments: a comparative before and after study. Ann R Coll Surg Engl 2012;94:388–394.
14 Wenner DM, Brody BA, Jarman AF, Kolman JM, Wray NP, Ashton CM: Do surgical trials meet the scientific standards for clinical trials? J Am Coll Surg 2012;215:722–730.
15 Rosenthal R, Kasenda B, Dell-Kuster S, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, Amstutz A, Bengough T, Meerpohl JJ, Stegert M, Tikkinen KA, Neumann I, Carrasco-Labra A, Faulhaber M, Mulla S, Mertz D, Akl EA, Bassler D, Busse JW, Ferreira-Gonzalez I, Lamontagne F, Nordmann A, Gloy V, Olu KK, Raatz H, Moja L, Ebrahim S, Schandelmaier S, Sun X, Vandvik PO, Johnston BC, Walter MA, Burnand B, Schwenkglenks M, Hemkens LG, Bucher HC, Guyatt GH, Briel M: Completion and publication rates of randomized controlled trials in surgery: an empirical study. Ann Surg 2014, Epub ahead of print.
16 Strobel O, Buchler MW: The problem of the poor control arm in surgical randomized controlled trials. Br J Surg 2013;100:172–173.
17 Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. The EC/IC Bypass Study Group. N Engl J Med 1985;313:1191–1200.
18 Cobb LA, Thomas GI, Dillard DH, Merendino KA, Bruce RA: An evaluation of internal-mammary-artery ligation by a double-blind technic. N Engl J Med 1959;260:1115–1118.
19 Edwards JP, Kelly EJ, Lin Y, Lenders T, Ghali WA, Graham AJ: Meta-analytic comparison of randomized and nonrandomized studies of breast cancer surgery. Can J Surg 2012;55:155–162.
20 Peinemann F, Tushabe DA, Kleijnen J: Using multiple types of studies in systematic reviews of health care interventions – a systematic review. PLoS One 2013;8:e85035.
21 Potter S, Mills N, Cawthorn SJ, Donovan J, Blazeby JM: Time to be BRAVE: is educating surgeons the key to unlocking the potential of randomised clinical trials in surgery? A qualitative study. Trials 2014;15:80.
22 Kaur G, Hutchison I, Mehanna H, Williamson P, Shaw R, Tudur Smith C: Barriers to recruitment for surgical trials in head and neck oncology: a survey of trial investigators. BMJ Open 2013;3:e002625.

23 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.
24 Thompson JE, Egger S, Bohm M, Haynes AM, Matthews J, Rasiah K, Stricker PD: Superior quality of life and improved surgical margins are achievable with robotic radical prostatectomy after a long learning curve: a prospective single-surgeon study of 1552 consecutive cases. Eur Urol 2014;65:521–531.
25 Le Morvan P, Stock B: Medical learning curves and the Kantian ideal. J Med Ethics 2005;31:513–518.
26 Hartgrink HH, van de Velde CJ, Putter H, Bonenkamp JJ, Klein Kranenbarg E, Songun I, Welvaart K, van Krieken JH, Meijer S, Plukker JT, van Elk PJ, Obertop H, Gouma DJ, van Lanschot JJ, Taat CW, de Graaf PW, von Meyenfeldt MF, Tilanus H, Sasako M: Extended lymph node dissection for gastric cancer: who may benefit? Final results of the randomized Dutch gastric cancer group trial. J Clin Oncol 2004;22:2069–2077.
27 Bruce J, Krukowski ZH, Al-Khairy G, Russell EM, Park KG: Systematic review of the definition and measurement of anastomotic leak after gastrointestinal surgery. Br J Surg 2001;88:1157–1168.
28 Blencowe NS, Strong S, McNair AG, Brookes ST, Crosby T, Griffin SM, Blazeby JM: Reporting of short-term clinical outcomes after esophagectomy: a systematic review. Ann Surg 2012;255:658–666.
29 Chan S, Bhandari M: The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. J Bone Joint Surg Am 2007;89:1970–1978.
30 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.
31 Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H, Kalske J, Jarvinen TL; Finnish Degenerative Meniscal Lesion Study G: Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. N Engl J Med 2013;369:2515–2524.
32 Mann M, Tendulkar A, Birger N, Howard C, Ratcliffe MB: National institutes of health funding for surgical research. Ann Surg 2008;247:217–221.
33 Ratcliffe MB, Howard C, Mann M, del Nido P: National Institutes of Health funding for cardiothoracic surgical research. J Thorac Cardiovasc Surg 2008;136:392–397; discussion 398–399.
34 McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D: Randomised trials in surgery: problems and possible solutions. BMJ 2002;324:1448–1451.
35 Jarman AF, Wray NP, Wenner DM, Ashton CM: Trials and tribulations: the professional development of surgical trialists. Am J Surg 2012;204:339–346.e5.
36 Challoner DR, Vodra WW: Medical devices and health – creating a new regulatory framework for moderate-risk devices. N Engl J Med 2011;365:977–979.
37 Campillo-Artero C: A full-fledged overhaul is needed for a risk and value-based regulation of medical devices in Europe. Health Policy 2013;113:38–44.
38 Concato J, Shah N, Horwitz RI: Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000;342:1887–1892.
39 Glasziou P, Chalmers I, Rawlins M, McCulloch P: When are randomised trials unnecessary? Picking signal from noise. BMJ 2007;334:349–351.
40 Lassen K, Hvarphiye A, Myrmel T: Randomised trials in surgery: the burden of evidence. Rev Recent Clin Trials 2012;7:244–248.
41 Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas WB: Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. BMC Med Res Methodol 2006;6:29.
42 Adamina M, Guller U, Weber WP, Oertli D: Propensity scores and the surgeon. Br J Surg 2006;93:389–394.
43 Brown CA, Lilford RJ: The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006; 6:54.
44 Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, Seiler CM, Balliol C, Altman DG, Aronson JK, Barkun JS, Campbell WB, Cook JA, Feldman LS, Flum DR, Glasziou P, Maddern GJ, Marshall JC, McCulloch P, Nicholl J, Strasberg SM, Meakins JL, Ashby D, Black N, Bunker J, Burton M, Campbell M, Chalkidou K, Chalmers I, de Leval M, Deeks J, Grant A, Gray M, Greenhalgh R, Jenicek M, Kehoe S, Lilford R, Littlejohns P, Loke Y, Madhock R, McPherson K, Rothwell P, Summerskill B, Taggart D, Tekkis P, Thompson M, Treasure T, Trohler U, Vandenbroucke J: Challenges in evaluating surgical innovation. Lancet 2009;374:1097–1104.
45 McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Nicholl J, Balliol C, Aronson JK, Barkun JS, Blazeby JM, Boutron IC, Campbell WB, Clavien PA, Cook JA, Ergina PL, Feldman LS, Flum DR, Maddern GJ, Nicholl J, Reeves BC, Seiler CM, Strasberg SM, Meakins JL, Ashby D, Black N, Bunker J, Burton M, Campbell M, Chalkidou K, Chalmers I, de Leval M, Deeks J, Ergina PL, Grant A, Gray M, Greenhalgh R, Jenicek M, Kehoe S, Lilford R, Littlejohns P, Loke Y, Madhock R, McPherson K, Meakins J, Rothwell P, Summerskill B, Taggart D, Tekkis P, Thompson M, Treasure T, Trohler U, Vandenbroucke J: No surgical innovation without evaluation: the IDEAL recommendations. Lancet 2009;374: 1105–1112.
46 Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A, IDEAL Group: IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. BMJ 2013;346:f2820.
47 Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG; IDEAL Group: IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. BMJ 2013;346:f3011.

Eur Surg Res 2014;53:95–107
DOI: 10.1159/000367606

48 McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK; IDEAL Group: IDEAL framework for surgical innovation 1: the idea and development stages. BMJ 2013;346:f3012.
49 The COMET (Core Outcome Measures in Effectiveness Trials) Initiative. http://www.comet-initiative.org/.
50 Macefield RC, Boulind CE, Blazeby JM: Selecting and measuring optimal outcomes for randomised controlled trials in surgery. Langenbecks Arch Surg 2014;399:263–272.
51 Antonescu I, Mueller CL, Fried GM, Vassiliou MC, Mayo NE, Feldman LS: Outcomes reported in high-impact surgical journals. Br J Surg 2014;101:582–589.
52 Waljee J, McGlinn EP, Sears ED, Chung KC: Patient expectations and patient-reported outcomes in surgery: a systematic review. Surgery 2014;155:799–808.
53 van der Pas MH, Haglind E, Cuesta MA, Furst A, Lacy AM, Hop WC, Bonjer HJ; Group COcLoORIS: Laparoscopic versus open surgery for rectal cancer (COLOR II): short-term outcomes of a randomised, phase 3 trial. Lancet Oncol 2013;14:210–218.
54 Cook JA, Ramsay CR, Fayers P: Statistical evaluation of learning curve effects in surgical trials. Clin Trials 2004;1:421–427.
55 Soreide K, Alderson D, Bergenfelz A, Beynon J, Connor S, Deckelbaum DL, Dejong CH, Earnshaw JJ, Kyamanywa P, Perez RO, Sakai Y, Winter DC; International Research Collaboration in Surgery ad-hoc working group: Strategies to improve clinical research in surgery through international collaboration. Lancet 2013;382:1140–1151.
56 Knebel P, Kuhn S, Ulrich AB, Buchler MW, Diener MK: The Study Centre of the German Surgical Society: current trials and results. Langenbecks Arch Surg 2012;397:611–618.
57 Royal College of Surgeons of England: The Rosetrees and the RCS clinical trials initiative. www.rcseng.ac.uk/surgeons/research/surgical-research/surgical-clinical-trials.
58 Neuman HB, Greenberg CC: Comparative effectiveness research: opportunities in surgical oncology. Semin Radiat Oncol 2014;24:43–48.
59 Bakaeen FG, Reda DJ, Gelijns AC, Cornwell L, Omer S, Al Jurdi R, Kougias P, Anaya D, Berger DH, Huang GD: Department of Veterans Affairs Cooperative Studies Program Network of Dedicated Enrollment Sites: implications for surgical trials. JAMA Surg 2014, Epub ahead of print.
60 Andres A, Toso C, Adam R, Barroso E, Hubert C, Capussotti L, Gerstel E, Roth A, Majno PE, Mentha G: A survival analysis of the liver-first reversed management of advanced simultaneous colorectal liver metastases: a LiverMetSurvey-based study. Ann Surg 2012;256:772–778; discussion 778–779.
61 Rice TW, Rusch VW, Apperson-Hansen C, Allen MS, Chen LQ, Hunter JG, Kesler KA, Law S, Lerut TE, Reed CE, Salo JA, Scott WJ, Swisher SG, Watson TJ, Blackstone EH: Worldwide esophageal cancer collaboration. Dis Esophagus 2009;22:1–8.
62 Breugom AJ, Boelens PG, van den Broek CB, Cervantes A, Van Cutsem E, Schmoll HJ, Valentini V, van de Velde CJ: Quality assurance in the treatment of colorectal cancer: the EURECCA initiative. Ann Oncol 2014;25;1485–1492.
63 Urquhart R, Grunfeld E, Jackson L, Sargeant J, Porter GA: Cross-disciplinary research in cancer: an opportunity to narrow the knowledge-practice gap. Curr Oncol 2013;20:e512–e521.
64 von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP; STROBE Initiative: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370:1453–1457.
65 European Commission press release, September 26, 2012. Safer, more effective and innovative medical devices. http://ec.europa.eu/health/medical-devices/files/revision_docs/pr_20120926_en.pdf.
66 A Delicate Balance: FDA and the Reform of the Medical Device Approval Process. http://www.fda.gov/NewsEvents/Testimony/ucm250709.htm.