OPINION

What can we learn from oncology surgical trials?

Serge Evrard, Pippa McKelvie-Sebileau, Cornelis van de Velde, Bernard Nordlinger and Graeme Poston

Abstract | Conducting high-quality prospective clinical trials in surgical oncology remains a challenge, and many seemingly well-designed trials lack this high quality because of inadequate recruitment accrual, lack of clinician interest, or evolution of treatment strategy during the many years over which such trials are conducted. In this Perspectives we examine some of the failures in published surgical oncology trials and discuss why they failed, and we make a critical assessment of the established prospective trial methodology in oncological practice (that is, phase 0, I, II, III and IV trials, and large prospective comparative audits) and how these methods might be used more effectively in future evaluation of cancer-surgery practice.

Evrard, S. *et al. Nat. Rev. Clin. Oncol*. advance online publication 20 October 2015; doi:10.1038/nrclinonc.2015.176

Introduction

Of the patients who are cured of cancer, 60% are cured by surgery alone; moreover, surgery has a major therapeutic role in curing the majority of the remaining 40%.[1,2] Another important consideration is the adoption of multidisciplinary management and population-based screening programmes. For example, if we consider breast cancer, 5-year survival rates have doubled from only 40% 50 years ago[3] to 87% today.[4] Likewise, 5-year survival rates for colorectal cancer (CRC), have more than doubled, from 23% 50 years ago to 57% at the present time.[5] Evidence to support the effectiveness of any given surgical procedure is nowhere near as strong as that for other parameters, such as overall survival and cure rates. Furthermore, outcomes from large prospective randomized trials rarely reflect surgical effectiveness,[6] and as such, provide a limited evidence base for such interventions; as a result, few patients are recruited into surgical trials. As a consequence of this situation, this weak evidence base impedes the adoption of new surgical procedures into clinical practice. A further problem is defining the research question and remit of a trial in surgical oncology: are such trials those that simply evaluate a single procedure—which might be considered too restrictive; or does this term encompass all cancer trials organized by a surgeon, such as those conducted by large consortia, for example, the National Surgical Adjuvant Breast and Bowel Project (NSABP)—which might be too expansive insofar as they address all aspects of treatment, other than just the surgical aspects. In reality, the majority of surgical trials examine the interface between a standardized procedure with or without a new intervention of potential benefit. What is important, however, is that, regardless of whether the trial addresses a comparison of two different surgical techniques or examines whether the addition of another treatment to the operation leads to better outcomes, it is essential that the trial design specifically standardizes the operative technique in order to be reproducible at all centres recruiting into the study.

A further problem that can affect the surgical quality control, which in turn can influence how patients are evaluated in trials of adjuvant therapy, is surgical specialty bias. One of the first published studies on this subject identified a common situation in which a 'nonspecialty' surgeon biopsied a metastatic lymph node and then immediately referred the patient into a randomized adjuvant therapy trial without completion lymphadenectomy.[7] When these patients had disease recurrence in the nodal basin within a few months, they were taken off the study as they were deemed a 'treatment failure', whereas their disease relapse could, in fact, have been avoided altogether if they had undergone a standardized surgical procedure (completion lymphadenectomy) at the time of initial lymph-node biopsy.[7]

Many of the challenges associated with conducting surgical trials include the need for a pragmatic study design, as well as the complexity of the interventions and the relative lack of experience among surgeons involved in clinical trials.[6] Many of these challenges have been addressed by the Balliol Collaboration, which led to the formation of the Idea, Development, Exploration, Assessment, Long-term study (IDEAL) framework.[8,9] Applying strict research methodology to clinical research that involves surgery is difficult, and many obstacles to trial design and execution are inherent to surgery.[6,10] Tumour heterogeneity and disease complexities make surgical procedures more difficult to study than patients enrolled in nonsurgical trials; this is because whereas drug dosage and timing can be titrated and adjusted according to patient circumstance in pharmacological studies, operative techniques and procedures, by definition, should be standardized, regardless of differences in patient-related factors.[10] As a result, the quality of reporting in the surgical literature is low compared with the available data for other medical specialties.[11,12] Thus, relevant research methodologies reside somewhere on the continuum between retrospective studies and randomized controlled trials (RCTs); as a consequence, RCTs are seriously underused.[12] In contrast to chemotherapy trials in oncology, where the intervention tested can be controlled to the precise details of dosage and timing, surgery is not a regimen containing a precise quantity of a specific treatment, but a set of acts performed in a (hopefully) predetermined sequence. Many studies outline the fact that one of the main prognostic factors regarding success of a surgical intervention is the decision-making process of the individual surgeon.[13] Completing a surgical procedure

Competing interests
The authors declare no competing interests.

is consequently something that cannot be standardized: it depends on the skill of the surgeon, decisions that will be made perioperatively, as well as the characteristics and preferences of the patient, such as anatomical variation, body-mass index, age, fitness, and so on. Each surgical procedure is thus unique.

The experience and technical ability of the surgeon is of great importance.[12] In 1991, McArdle and Hole[14] commented following an audit of CRC surgery among 13 surgeons in Glasgow, Scotland, that showed wide variations in all metrics of outcome. They commented that "some surgeons perform less than optimal surgery. Some surgeons are less competent technically than their colleagues; and some fail to supervise surgeons adequately during training. If by more meticulous attention to detail, the results of surgery could be improved, and our results suggest that this would not be difficult, the impact on survival might be greater than that of any of the adjuvant therapies currently under study".[14] Therefore, maintaining quality assurance, which is defined as the complete set of systematic actions that is required to achieve a treatment result that meets a certain standard agreed at the outset by the investigators (published within the study registration), is essential when designing trials in surgical oncology, and adjusting for confounding variables, such as the quality of operative surgery between recruiting centres and individual surgeons, and accounting for the possibility of a new procedure emerging during the duration of the study is critical.

Surgical trials are different

The application of evidence-based care to surgery has, in general, improved over the past decade,[5,15] but surgical interventions remain less likely to be investigated using full-scale and well-designed RCTs than other therapies.[5,15] A number of obstacles to conducting RCTs and other forms of clinical research in operative surgery exist. Firstly, the difficulties in standardization of the procedures and quality assurance can hinder reproducibility and external validity. The term 'surgery' comes from the Greek word *kheirourgía,* meaning manual activities. Surgery is a set of manual acts performed, ideally under the same circumstances, to treat each patient with a particular condition. The successful completion of a surgical procedure is consequently something that cannot be easily standardized. Moreover, the definition and the execution of a procedure can vary between centres: an example is the Dutch trial of D1 (removal of regional perigastric nodes) versus D2 (extended lymphadenectomy to include level 1 and level 2 regional nodes) lymphadenectomy during gastrectomy for gastric cancer;[16] a quality-control review by an expert Japanese surgeon of the procedures revealed that some so-called D2 procedures performed in the Dutch trial were actually closer to a D1 procedure.[17] This inability to define the types of surgery can confound results and hinder interpretation of outcome data.

Achieving good quality control for a procedure in a surgical study is a real challenge that greatly affects the quality of the conclusions that can be drawn. An example is the trial comparing postoperative chemoradiotherapy to surgery alone for gastric cancer, on the basis of the results of this study it was concluded that postoperative chemoradiotherapy after gastrectomy produced better results than gastrectomy alone.[18] In reality, the quality of the surgery was so poor (D0 and D1 lymphadenectomies) that postoperative chemoradiotherapy was simply compensating for poor surgery.[19,20]

Difficulties in patient accrual are a further obstacle to clinical trials in operative surgery; at least one in five surgical RCTs is discontinued early, mostly because of poor recruitment.[21] Accrual to surgical trials is more difficult than for other medical specialties because surgeons' reputations are difficult to establish and can be challenged by the methods of randomization. Although not unique to surgical oncology, surgical reputations in this setting might be harder to establish than in other surgical specialties (such as orthopaedics or cardiovascular surgery), the outcomes of which can be more predictable. In addition, most surgical trials address specific questions: the more specific the question, the more restrictive the criteria for inclusion, and thus fewer patients are eligible to be recruited to the study. The specificity of the issue addressed inversely affects the feasibility of the study. This fact can mean that clinical trials never get underway due to foreseen accrual difficulties, or other challenges that can be encountered during the study, resulting in trials being stopped early or downgraded—in the case of a recent intended randomized controlled trials with an enrollment target of more than 500 patients, to a 'pilot' randomized trial comprising only 100 patients.[22]

Another example of accrual and recruitment problems is illustrated by the conversion of the CLOCC Trial (EORTC 40004), which was initially designed as a phase III trial, into a randomized phase II trial.[23] The CLOCC trial was an interventional study that had difficulties in adequate patient recruitment and changes in clinical practice during recruitment that hindered the trial analysis. The study was designed in the late 1990s to test the hypothesis that addition of radiofrequency ablation (RFA) to systemic chemotherapy (FOLFOX4 or FOLFOX6 regimens, comprising 5-fluorouracil, folinic acid and oxaliplatin) would increase overall survival from 15 months (the median overall survival for FOLFOX therapy) to 30 months in patients with nonresectable liver-limited colorectal liver metastases (CRLM). The trial intended to recruit 400 patients from across Europe; however, radiofrequency ablation was widely available, which restricted the ability to recruit patients on to the trial, therefore, the trial was closed with only 119 patients enrolled. The time interval between the first-patient recruitment and trial reporting was almost 10 years, and during this time, the standard-of-care for CRLM changed from single first-line FOLFOX therapy to multiple lines of therapy after each instance of disease progression, with the addition of biologic agents, such as bevacizumab. The interim analysis at 1 year demonstrated a significant 20% ($P = 0.027$) improvement in progression free survival (PFS), the secondary end point of the study, in the experimental RFA arm compared with the standard-of-care.[23] As a result of the evolution of the standard-of-care, the median overall survival in both arms was similar at 30 months, which was the final follow-up duration of the trial. Despite a trend towards an overall survival benefit in the experimental arm after 5 years, the number of patients alive beyond 5 years (less than 10 patients in total at the time of reporting) was not sufficient to make meaningful statistical interpretation.[23] Nevertheless, the authors have now reported a significant improvement in overall survival for the experimental arm (32% overall survival for the addition of RFA compared to 5% for the chemotherapy arm, $P = 0.01$) at a median of 10 years after randomization.[24]

A further challenge to trial accrual is the physical and financial costs to the patient of travelling to a tertiary medical centre where a trial involving complex surgery might be conducted. This situation can also affect patient amenability to accepting

subsequent therapy, and lead to 'referral bias' in the recruitment process.[25] The lack of clinical equipoise can be another obstacle to randomization. Surgeons are often convinced that what they do is the best for their patients, when other ways to achieve similar or better outcomes might actually exist. Awareness of this observation is the foundation for surgical equipoise; however, the lack of equipoise might also come from the patient, who comes to the surgeon with their own ideas about the best technique, and so refuses randomization to a trial. As such, prospective evaluation of emerging techniques can be influenced and perverted by popularization in the internet and media.[8] An example of the problems posed by perceived equipoise was patient recruitment to the UK FACS study, which examined the benefits of different intensities of follow up following apparently curative resection of primary CRC, essentially examining the cost effectiveness of minimal primary-care review through to intensive hospital-based follow up.[26] Recruitment to this trial required investigators and patients to accept the equipoise that all methods of follow up were equal; however, the study successfully recruited and demonstrated the survival benefits and cost-effectiveness of more-intensive follow up. As a result, more patients who developed disease recurrence were detected at an earlier stage of recurrence when undergoing more-intensive follow up, and were, therefore, more amenable to more-effective therapeutic interventions at an earlier stage, when their disease was more-effectively treatable.

Administration associated with carrying out cancer surgery trials can be challenging. The American College of Surgeons created an oncology group (ACOSOG) with the intention of performing RCTs for surgical oncology procedures. The ACOSOG Z0011 study, although underpowered, was ultimately practice changing for breast surgeons.[27–29] This study compared immediate versus delayed axillary-lymph-node clearance in women found to have a tumour-positive sentinel lymph node during breast-cancer surgery. All oncological outcomes (disease-free and overall survival) were not dissimilar between the two groups, whereas operative complications and postoperative morbidity were higher following immediate axillary clearance.[27–29] Unfortunately, the ACOSOG no longer exists. This group essentially disbanded during further consolidation of larger oncology groups, although all current ACOSOG protocols and documents remain available via the Alliance for Clinical Trials website.[30] Surgeons continue to be well represented within the Alliance, and one ongoing study, Z4099, has a surgeon credentialing checklist that is an example of the goals that pervaded ACOSOG (Box 1).[30,31]

Box 1 | Surgeon credentialing for the ACOSOG Z4099/RTOG 1021 trial[31]

The ACOSOG Z4099/RTOG 1021 trial is a randomized phase III study that is comparing sublobar resection (± brachytherapy) and stereotactic body radiation therapy in the treatment of patients with high-risk stage I NSCLC.[30]

- In this trial, credentialing of all surgeons will be conducted by the study chair or designee
- Participating surgeons must complete and submit the Z4099 Surgeon Credentialing Checklist available on the Z4099 page of the ACOSOG website[30] before registering a patient

Surgeons must be either a member of the General Thoracic Surgery Club or a board-certified cardiothoracic surgeon with >50% of surgery practice devoted to general thoracic surgery. Criteria for membership of the General Thoracic Surgery Club include:

- Surgeons who have obtained specialty certification in thoracic surgery by the American Board of Thoracic Surgery or the Royal College of Surgeons, or other official certifying organization
- Surgeons who have been in practice for a minimum of 2 years beyond the completion of formal training in thoracic surgery, and devote at least 50% of their practice to general thoracic surgery
- Surgeons whose list of all operations performed in the year before application has been certified by the chief(s) of surgery at their institution(s)

Surgeons who do not meet the above criteria must submit the following for review by the study chair:

- Case list of operative experience for the previous year
- Operative and pathology reports for five sublobar resection procedures done during the previous year

Abbreviations: ACOSOG, American College of Surgeons Oncology Group; NSCLC, non-small-cell lung cancer; RTOG, Radiation Therapy Oncology Group.

Finally, funding issues often confound research objectives. Surgical endeavours in modern medicine are often not sufficiently recognized by funding agencies.[11] The proportion of public funding devoted to surgically oriented clinical research is very small (<5%) compared with that allocated to 'medical' trials.[32,33] The study design in RCTs of surgical intervention is described insufficiently in more than 60% of published papers;[34] thus, efforts are required to improve study design, as well as reporting methods. Pharmaceutical companies do support trials addressing the addition of perioperative systemic therapies to improve (mostly progression-free) survival after apparent curative cancer surgery, and increasingly trials address converting unresectable tumours to resectable tumours with curative intent. An example of how the quality of the outcome of such an intervention can be calculated lies within the outcome of NICE Technology Appraisal (TA) 176, which recommends the use of the anti-EGFR antibody cetuximab with systemic chemotherapy to treat patients with unresectable *KRAS*-wild-type CRLM, with the intention of conversion to resectability with curative intent.[35] The health-economics modelling showed that, if this end point could be achieved, the incremental cost-effectiveness ratio (ICER) per quality-adjusted life year (QALY) fell dramatically from over £50,000 to under £30,000, and so the intervention was considered cost-effective. In general, however, companies are not interested in supporting phase I trials examining a new surgical innovation, and unfortunately, most of the instrument manufacturers have little history of bringing such products to market through a rigorous clinical trial methodology.

A pragmatic approach

The future perspectives for trials in surgical oncology have been recently reviewed in this journal,[36] although we propose a pragmatic approach to addressing the deficiencies in our current approach to conducting trials in this field. Various classification hierarchies of levels of evidence have been proposed with the best evidence derived from the highest levels leading to enhanced care for patients, and improved standards for surgeons and health-care institutions. The Canadian Task Force on the Periodic Health Examination proposed a classification system from I to III in 1979.[37] This levels-of-evidence system was further developed and expanded upon by Sackett in 1989 for antithrombotic agents (Box 2).[38] Both classifications agree that for therapeutic studies, only RCTs can provide level I evidence. However, some of the most important published surgical research is based on retrospective studies, including studies reported in high impact-factor journals,[39–41] often

Box 2 | Levels and types of evidence from the Canadian Task Force*

Canadian Task Force on the Periodic Health Examination
- Level I: At least one randomized controlled trial with proper randomization
- Level II.1: Well-designed cohort or case-control study
- Level II.2: Time series comparisons or dramatic results from uncontrolled studies
- Level III: Expert opinions

Sackett levels of evidence
- Level I: Large randomized controlled trials with clear-cut results
- Level II: Small randomized controlled trials with unclear results
- Level III: Cohort and case–control studies
- Level IV: Historical cohort or case–control studies
- Level V: Case series, studies with no controls

*Based on information from the Canadian Task Force on the Periodic Health Examination[37] and Sackett (1989).[38]

because randomizing patients between two very diverse treatment arms is impossible, although this limitation is not always explicitly stated as the reason for not undertaking a prospective RCT.[40,42]

The IDEAL recommendations[6,9,43] for the assessment of surgical outcomes, based on a five-stage description of the surgical development process, are a good inventory of the difficulties of managing and organizing clinical research in surgery. These recommendations include registration as a professional duty (anonymously, if necessary, when outcomes are adverse) of reports of new techniques, and case-series studies to be replaced by prospective development studies for early technical modifications and by prospective research databases for later pre-trial evaluation.[44,45] Much can be learnt from the use of prospective registries for the introduction of innovative procedures in other aspects of surgery, such as the introduction of natural orifice transluminal endoscopic surgery (NOTES) for benign intra-abdominal conditions,[44] and in oncology, the introduction of associating liver partition and portal vein ligation (the ALPPS procedure), the modified 2-stage hepatectomy procedure for liver tumours—for which the registry demonstrated increased perioperative morbidity and mortality in older patients compared to conventional liver resection procedures.[45] Protocols for surgical studies should be registered publicly, and the statistical process to control for the techniques considered in both early and late assessment. Lastly, randomized trials should be used whenever possible to investigate efficacy, but adequate pretrial data are essential to enable power calculations, to clarify the definition and indications of the intervention, and to develop quality measures. However, the scope for such randomized trials remains fairly theoretical, and the principals are difficult to apply within the time and funding constraints of applying surgical research within real practice.

Critique of surgical RCTs

Surgical innovation—initially reported mainly through case studies, small retrospective series, or the media—sometimes seems to be so 'obviously' relevant that such innovations are adopted into routine practice without comment, before finally being undermined years later by an RCT. One such example is coronary-artery stenting for stable coronary-artery disease.[46] Phase III trials are rare in all aspects of surgery, representing only 8% of publications relating to this modality in 2006.[47] Retrospective studies, which are cheaper, faster and easier to carry out, have a role in clinical research in surgery, but RCTs are the ideal approach, and clear and accurate reporting of RCT results is needed to guide evidence-based medicine. Despite being the 'gold standard', the RCT is not the only valid method to evaluate a new treatment. Some experimental studies that compare estimations of treatment effects using either an RCT or nonrandomized trials have not shown a definitive superiority of RCTs.[47,48] RCTs are limited by their lack of generalizability: they require strict inclusion and exclusion criteria, and major difficulties in trial accrual can hinder surgical RCTs owing to the rarity of patients meeting inclusion criteria. Furthermore, selection and publication bias can influence the study conclusions. This situation highlights the issue that outcomes for patients recruited into RCTs are not representative of those for patients receiving surgical care in the 'real' world. Patients not on RCTs might be better assessed by prospective population-based research, which would ideally include the total population under study, an example of which is the European Registration of Cancer Care comparative outcome audit (EURECCA).[49] RCTs are necessary, however, and they should be preceded by other trial phases, to firmly establish criteria for sample-size calculation and expected outcomes.

The Consolidated Standards of Reporting Trials (CONSORT) statement[50,51] was extended in 2008 to randomized trials of nonpharmacologic treatment (CONSORT-NPT, Consolidated Standards of Reporting Trials for Non Pharmacological Trials),[52] and incorporated elements such as the complexity of the intervention, expertise of the care provider, and difficulties of blinding. Nagendran and colleagues[53] have reported worse adherence to the CONSORT-NPT extension than to the original general CONSORT recommendations, possibly because there has been less time to adapt reporting practices and journal policy to the CONSORT-NPT guidelines published in 2008—more than 10 years after the original CONSORT statement. Indeed, the investigators reported limited awareness of the CONSORT-NPT statement by authors, peer-reviewers, and editors. In a recent effort to increase adherence to these guidelines, the checklist was simplified to a 10-item reduced checklist.[54]

The major risk of fewer and poorer quality surgical RCTs is to impinge on quality assurance. For example, in 2002, a randomized trial of laparoscopy-assisted colectomy (LAC) versus open colectomy for the treatment of nonmetastatic colon cancer was reported.[55] The investigators concluded that LAC was more effective than open colectomy in this setting, in terms of morbidity, duration of hospital stay, tumour recurrence, and also cancer-related survival. This trial was the first randomized study published on this topic, and the remit of LAC at that time was a pressing health-economics issue; thus, the result was considered practice changing. All the requirements for publication were met except one, the quality assurance of the methodology. Briefly, the trial was underpowered and multi-biased, and whether the randomization design was written before or after completion of patient accrual was not clear.[56] To prevent such criticism, future trials should only be considered prospective if the study is registered before starting the first inclusion. These results of this LAC trial were never confirmed in subsequent publications, and the two methods of colectomy are now considered to be equivalent.[57] It is easy to understand the danger of publishing such a 'randomized trial' that does not adhere to the CONSORT criteria. As prospective RCTs are rare, successful completion is a major achievement, and consequently rigorous control of quality assurance of such trials is mandatory. Nonetheless, the reason why surgical RCTs often do not conform

to CONSORT criteria is mainly because phase III frameworks are not designed for evaluation of nonpharmacological interventions. Indeed, all the reasons why such prospective clinical research is more difficult in surgery than in medicine are present and exacerbated in the phase III trial concept: lack of equipoise; difficulty in patient accrual; standardization and control of quality of the procedures; and finally the cost and the time required (usually in years) to recruit and reach the primary end point (usually at least 3-year disease-free survival). Such an example is the Dutch D1 versus D2 lymphadenectomy trial in patients with gastric cancer, in which the real oncological benefit of the more-radical lymphadenectomy wasn't demonstrated until analysis of the 15-year overall survival follow-up data.[17]

Moreover, as surgical randomized trials have difficulties in adhering to the CONSORT criteria; the risk is that after several years of efforts, they will be criticized on publication, considered as biased, and finally discarded. A good example of such criticism followed the publication of the ASTEC (A Study in the Treatment of Endometrial Cancer) surgical trial,[58] which investigated whether pelvic lymphadenectomy could improve the survival of women with grade 1 endometrial cancer. In total, 85 centres in four countries and 1,408 women participated. The results showed no evidence of benefit in terms of overall survival or recurrence-free survival. Some of the authors of the trial publication subsequently criticized the study, outlining that the number of the retrieved lymph nodes was too small, the inclusion rate of low-risk patients was too high, and periaortic lymph nodes were left *in situ*.[59] Ultimately, all authors concluded, more or less, that they would continue to undertake such lymphadenectomies for women with this condition.

A similar controversy[60] arose following the publication of the New EPOC trial that randomized patients with resectable CRLM to perioperative systemic (FOLFOX) chemotherapy with or without cetuximab.[60] This trial was a UK-based study in which any liver surgery centre could participate, but no real-time quality assurance mechanism was in place to assess the standard of surgery performed. Ultimately, the trial was closed early because of what seemed to be a reduction in progression-free survival (PFS) among patients receiving FOLFOX plus cetuximab. The trial investigators intended to recruit patients with easily resectable metastases, but only 33% of patients were reported to have R0 (curative) resections. The remaining two-third of patients had either positive resection margins or no trace of tumour in the operative specimen, because the trial was designed to achieve R0 resection, and if this is not achieved, this reflects on surgical quality.[61] When surprising effects are observed in clinical trials, with no clear underlying biological explanation, the quality of the intervention across participating centres is of paramount importance to enable the correct interpretation of the clinical findings.

In summary, two major potential drawbacks face phase III trials that test the benefit of a surgical intervention: the first is the risk of a high level of investment of time and effort required in a study, which several years later will be judged as having failed to demonstrate anything; the second possible drawback is the more-serious risk of establishing a flawed hypothesis as the truth. Other alternative prospective designs are often more feasible, for which compliance with statistical principles, such as minimum sample size, is more likely. These study designs have a crucial role in clinical research in surgical oncology.

Alternative prospective designs

If conducting prospective phase III trials is difficult in surgical oncology, what are the possible alternatives? A number of possible prospective nonrandomized evaluation frameworks are available that might be appropriate to enhance the scientific level of reporting in surgical oncology. We discuss these alternatives in the next sections and provide a framework for improved surgical trial outcomes.

Phase 0

Phase 0 trials are also referred to as exploratory investigational new-drug studies, or exploratory trials. The aim of these very early phase trials is to demonstrate biological efficacy of the agent before the first phase I trial. In general, these trials are used to help speed up and streamline the drug-approval process by establishing very early on in this process whether the agent behaves in human subjects as was expected from preclinical studies. The principle is to give a small quantity of a new drug (1/100 of the proposed effective dose) to a small group of patients (fewer than 10) over a short period (over fewer than 7 days).

These studies are relatively difficult to organize, as they have provide no individual benefit to patients and usually require invasive biopsies, which is why surgeons have a major role in such trials. One possibility is to take advantage of routine surgery to provide tumour tissue to assess drug response without the need for additional invasive biopsies. For example, early stage colon cancer that is histologically confirmed by colonoscopy is usually treated by surgical resection after a short interval of 2–3 weeks, during which a new pharmacological agent can be administered and its effect on the tumour thereby evaluated. Phase 0 trials, therefore, offer an excellent link between surgeons and translational biologists. The surgeon should take the initiative in identifying suitable procedures (a standard elective cancer operation that can be organized without being disturbed by a short period of drug administration) and collaborate with the translational biologist in preliminary testing of a new agent for the first time outside of an animal model.

Phase I

Phase I trials are designed to assess the safety of a new intervention in a small group of patients; however, these trials do not exist for surgical procedures, and no framework exists for surgical innovation. For example, mandatory CE marking (CE mark, or formerly EC mark, is a mandatory conformity marking for certain products sold within the European Economic Area (EEA) since 1985. The CE marking is also found on products sold outside the EEA that are manufactured in, or designed to be sold in, the EEA, for a new surgical device can attest to the quality and safety of the device production, and compliance with industry standards, but cannot guarantee the legitimacy of indications for its use in humans. All these requirements have to be established after the device has reached the market. For example, when the first radiofrequency tumour ablation generators for liver tumours were implemented in practice at the end of the 20th century, no recognized guidelines were available on how to use radiofrequency for this indication—for which lesions, which size, and so on—and the first prospective evaluations were not published until 2012.[23,62]

Even greater concern surrounds the adoption of new procedures. A major issue in surgical research is innovation that seems so obvious that the approach is adopted immediately as a validated practice, although, strictly, it has never passed the experimental and validation phases of development. What official regulatory authority is needed for first-in-human procedures? The IDEAL recommendations advocate open-access registries that record technical details

Box 3 | The IDEAL recommendations for surgical trial design[6,43]

Assessment of surgical innovations depend on the following factors:
- Operator
- Team
- Setting
- Learning curves
- Quality variation
- Perception of equipoise

Reports of new techniques should be registered as a professional duty, anonymously if necessary, when outcomes are adverse

Encourage the widespread use of prospective databases and registries

Case series studies replaced by either prospective development studies for early technical modifications or prospective research databases for later pre-trial evaluation

Study protocols should be publicly registered

Statistical process for the control techniques useful in both early and late assessment

Randomized trials should be used whenever possible to investigate efficacy, but adequate pre-trial data are essential to allow power calculations, and clarify the definition and indications for the intervention as well as develop quality measures

Difficulties in doing randomized trials addressed by measures to evaluate learning curves, and alleviate equipoise problems

Alternative prospective designs (e.g. interrupted time series studies) should be used when randomized trials not feasible

Established procedures should be monitored with prospective databases to analyse outcome variations, and identify late and rare events

Achievement of improved design, conduct, and reporting of surgical research will need concerted action by: editors, funders of healthcare and research, regulatory bodies, and professional societies

Abbreviation: IDEAL, Idea, Development, Exploration, Assessment, Long-term study framework.

and pitfalls.[63] As such, health regulatory authorities need to establish a methodological framework for evaluating new surgical devices and techniques, and their indications, before widespread adoption—similar to the market approval of new drugs.

Phase II

Phase II trials are intended to establish the potential efficacy of a new treatment. The phase II framework is quite easily adaptable for surgery, but is largely underused in surgical trials. The principle is to define *a priori* a threshold for efficacy and non-efficacy. As short-term efficacy is often the focus of the primary end point, in the surgical setting, it will favour the effectiveness of response to the intervention rather than overall survival. If an immediate response rate cannot be used, it can be replaced by disease-free survival. The number of patients required is based on power calculations, but is usually around 30–50, with the possibility of using several optimal designs, and potentially reducing the required number of patients by introducing an intermediate level of analysis.[64] Twin phase II trials (concurrent linked trials) can be run by randomization, such as conducted for the CLOCC trial[23,24] or in parallel without randomization, such as the ongoing Cascador study involving delayed coloanal anastomosis for medium and lower rectal cancer treatment.[65] This corresponds to the IDEAL stage 3 evaluation or nonrandomized controlled trials (Box 3).[42] In this setting, which differs from RCTs only by the absence of randomization and blinding, comparisons between the two arms are not possible; however, twin phase II trials enable the assessment of treatment effect simultaneously in both groups. The phase II design is generally associated with a high false-positive rate (often around 10% or more), which is responsible for its poor reputation in medical oncology. However, a well-conducted phase II trial has more value in surgery than in medicine, owing to the inherent difficulties of accumulating high-quality prospective data.

Phase IV

Phase IV trials are usually designed as post-marketing surveillance studies to monitor the risks, benefits, and optimal use of a treatment once it has been approved for use. The problem with respect to surgical practice is that the preliminary phases in the introduction of a new surgical strategy are often skipped, and so phase IV studies become the first attempt at a prospective methodological evaluation. As an example, CRLM involving both lobes of the liver has been routinely resected all over the world for more than 30 years using one-stage or two-stage procedures, with or without portal vein embolization and tumour ablation, without any prospective evaluation until 2012, when two phase II trials where published to evaluate the use of combined resection and ablation for advanced-stage cancers.[23,62]

When a procedure has been accepted as routine clinical practice and it is too late for phase II or III trials to be conducted, carrying out a phase IV study is always possible, especially when health-economics concerns exist. One such proposed outcome study follows the recent strategic collaboration between the European Society of Surgical Oncology (ESSO) and the EORTC: the CLIMB (Colorectal Liver Metastasis Database) study (NCT02218801),[66] which is a prospective multicentre database recording advanced cases of CRLM that are considered initially unresectable and require chemotherapy with the intention to convert to resectability, or are operable upfront (by combined resection and ablation). The primary end points will be operative morbidity and disease-free survival, but a programme to improve the quality of surgery is also scheduled. After the first 100 cases, further specific metrics will be identified to become targets to define a policy for outcome improvement.

Quality assurance—assessing benefits

Many trials of adjuvant and neoadjuvant systemic therapies have been conducted by medical oncologists preceding or following apparently curative resections of primary solid cancers. Few of these trials have been surgically driven, however, an example of one such surgically driven study that confirmed the improvement possible in reduced local recurrence rates by the addition of neoadjuvant therapy, is the Swedish radiotherapy trial for rectal cancer surgery.[67] However, contrary to the Swedish trial,[66] the Dutch total mesorectal excision (TME) trial[67] standardized the surgical procedure of TME and randomized patients to either surgery alone or preoperative 25 Gy radiotherapy (delivered as 5 × 5 Gy fractions) followed by surgery. The addition of the preoperative radiotherapy reduced 5-year local recurrence rates following TME surgery from 10.9% to 5.6% in the Dutch trial.[68] This improvement was also noted when the data from this study were subsequently pooled with similar studies conducted in the Netherlands, Poland, and by the EORTC.[69]

Another such surgically driven trial was the EPOC study (EORTC 40983), which tested the hypothesis that the addition of perioperative systemic chemotherapy (FOLFOX4) to surgery would improve

3-year PFS, compared with surgery alone, in patients with easily resectable CRLM.[70] In retrospect, the study suffered a number of flaws that included recruiting 60 of 360 patients into the randomization who were subsequently found to be ineligible (understaged inoperable disease, unresectable extra-hepatic disease, or didn't actually have CRLM). As such, the intention-to-treat final analysis showed the difference in PFS at 3 years was not significant, but a subset analysis of those patients who did have CRLM and were resected (300 of those randomized), showed a statistically significant benefit (9% improvement in PFS, $P = 0.035$) that favoured the chemotherapy arm.[70]

Outcome improvements?

Good data demonstrate that in the aftermath of large-scale national studies testing more-appropriate surgical strategies compared to conventional techniques, if the newer technique achieves better overall oncological outcomes in terms of local disease control and overall survival, then they are rapidly taken up by the surgical community. Examples include the Dutch trial of TME for rectal cancer,[70] following the publication of which, national outcome figures in patients with this disease improved considerably,[72,73] and also following the Dutch D1 versus D2 lymphadenectomy trial for gastric cancer,[16,17] oncological outcomes of which improved in the Netherlands thereafter.[74] However, much remains to be achieved. Despite clear evidence of the benefits shown by surgical trials of operating in the appropriate oncological anatomical plane for CRC, only 32% of colon cancer operations are being conducted in the appropriate mesocolic plane,[75] and only 53% of rectal cancer patients undergo an appropriate TME, with 47% of patients continuing to remain at risk of cancer-threatened circumferential resection margins.[76,77]

Conclusions

No ideal design exists for an optimal clinical trial in cancer surgery. Tradeoffs between the exigency of high level of scientific evidence and study feasibility are inevitable. Surgical oncologists must be convinced of the necessity to run prospective trials or studies of differing types depending on the issues to be addressed, and accept that in many cases randomization will not be possible, in which case large-scale prospective comparative audit remains a powerful instrument.[40] Recognizing this urgent need, EORTC and ESSO initiated such a collaboration that will use each organization's best capabilities to advance surgical oncology research in Europe. This programme is 'SURCARE: An EORTC–ESSO Initiative for High Quality Standards in Prospective Surgical Clinical Research'.[65] SURCARE is a comprehensive surgical task force, which hopefully will develop prospective, high-quality research that focuses on the outcomes most relevant to our patients. Another issue to be addressed is the complete lack of trials addressing issues of cancer surgery in the elderly. In parallel, journal editors and peer reviewers have a responsibility to endorse guidelines for the scientific quality of surgical research, such as the CONSORT-NPT statement or the IDEAL guidelines, to improve the level of the evidence-base in surgical oncology. Publication criteria should also be broadened to include other prospective research methodologies and an end to the monopoly of the RCT in surgical oncology. If this is achieved, then major advances in surgical outcome can be expected in the future.

*Institut Bergonie, 229 cours de l'Argonne, Bordeaux 33000, France (**S.E., P.M.-S.**). Department of Surgery, Leiden University Medical Centre, Albinusdreef 2, 2300 RC, Leiden, Netherlands (**C.v.d.V.**). Hopital Ambroise Pare, Department of Surgery, 9 Avenue Charles de Gaulle, 92104 Boulogne-Billancourt, France (**B.N.**). Department of Surgery, School of Translational Studies, University of Liverpool, and Department of Surgery, Aintree University Hospital, Liverpool L9 7AL, UK (**G.P.**).*
Correspondence to: G.P.
graeme.poston@aintree.nhs.uk

1. Wyld, L., Audisio, R. A. & Poston, G. J. The evolution of cancer surgery and future perspectives. *Nat. Rev. Clin. Oncol.* **12**, 115–124 (2015).
2. Poston, G. J. in *Textbook of Surgical Oncology* (eds Poston, G. J., Beauchamp, R. D. & Ruers, T.) 1–4 (Informa Healthcare, 2007).
3. Philip, J. F. Results of treatment of breast cancer at Aberdeen Royal Infirmary 1940–55. *Br. Med. J.* **1**, 323–331 (1967).
4. Cancer Research UK. *Cancer Statistics, Key Facts: All Cancers Combined, February 2015* [online], http://publications.cancerresearchuk.org/downloads/Product/CS_KF_ALLCANCERS.pdf (2015).
5. McCulloch, P., Taylor, I., Sasako, M., Lovett, B. & Griffin, D. Randomised trials in surgery: problems and possible solutions. *BMJ* **324**, 1448–1451 (2002).
6. Ergina, P. L. *et al.* Challenges in evaluating surgical innovation. *Lancet* **374**, 1097–1104 (2009).
7. Balch, C. M., Durant, J. R. & Bartolucci, A. A. The impact of surgical quality control in multi-institutional group trials involving adjuvant cancer treatments. *Ann. Surg.* **198**, 164–167 (1983).
8. [No authors listed] Surgical research: the reality and the IDEAL. *Lancet* **374**, 1037 (2009).
9. McCulloch, P. *et al.* No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* **374**, 1105–1112 (2009).
10. Solomon, M. J. & McLeod, R. S. Clinical studies in surgical journals—have we improved? *Dis. Colon Rectum* **36**, 43–48 (1993).
11. Weil, R. J. The future of surgical research. *PLoS Med.* **1**, e13 (2004).
12. Panesar, S. S., Thakrar, R., Athanasiou, T. & Sheikh, A. Comparison of reports of randomized controlled trials and systematic reviews in surgical journals: literature review. *J. R. Soc. Med.* **99**, 470–472 (2006).
13. Martling, A., Cedermark, B., Johansson, H., Rutqvist, L. E. & Holm, T. The surgeon as a prognostic factor after the introduction of total mesorectal excision in the treatment of rectal cancer. *Br. J. Surg.* **89**, 1008–1013 (2002).
14. McArdle, C. S. & Hole, D. Impact of variability among surgeons on postoperative morbidity and mortality and ultimate survival. *BMJ* **302**, 1501–1505 (1991).
15. Farrokhyar, F. *et al.* Randomized controlled trials of surgical interventions. *Ann. Surg.* **251**, 409–416 (2010).
16. Bonenkamp, J. J. *et al.* Extended lymph-node dissection for gastric cancer. *N. Engl. J. Med.* **340**, 908–914 (1999).
17. Songun, I., Putter, H., Kranenbarg, E. M., Sasako, M. & van de Velde, C. J. Surgical treatment of gastric cancer: 15-year follow-up results of the randomised nationwide Dutch D1D2 trial. *Lancet Oncol.* **11**, 439–49 (2010).
18. Macdonald, J. S. *et al.* Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. *N. Engl. J. Med.* **345**, 725–730 (2001).
19. Abraham, I., Dhar, P. & Praseedom, R. K. Adjuvant chemoradiotherapy for gastric cancer. *N. Engl. J. Med.* **346**, 210–211 (2002).
20. Cuschieri, A. Does chemoradiotherapy after intended curative surgery increase survival of gastric cancer patients? *Gut* **50**, 751 (2002).
21. Chapman, S. J. *et al.* Discontinuation and non-publication of surgical randomised controlled trials: observational study. *BMJ* **349**, g6870 (2014).
22. Denost, Q. *et al.* Perineal transanal approach: a new standard for laparoscopic sphincter-saving resection in low rectal cancer, a randomized trial. *Ann. Surg.* **260**, 993–999 (2014).
23. Ruers, T. *et al.* Radiofrequency ablation combined with systemic treatment versus systemic treatment alone in patients with non-resectable colorectal liver metastases: a randomized EORTC Intergroup phase II study (EORTC 40004). *Ann. Oncol.* **23**, 2619–2626 (2012).
24. Ruers, T. *et al.* Radiofrequency ablation (RFA) combined with chemotherapy for unresectable colorectal liver metastases (CRC LM): long-term survival results of a randomized phase II study of the EORTC-NCRI CCSG-ALM Intergroup 40004 (CLOCC) [abstract]. *J. Clin. Oncol.* **33** (Suppl.), a3501 (2015).
25. Lamont, E. B. *et al.* Is patient travel distance associated with survival on phase II clinical trials in oncology? *J. Natl Cancer Inst.* **95**, 1370–1375 (2003).
26. Pugh, S. A. *et al.* Site and stage of colorectal cancer influence the likelihood and distribution of disease recurrence and postrecurrence survival: data from the FACS randomized controlled trial. *Ann. Surg.* http://dx.doi.org/10.1097/SLA.0000000000001351 (2015).

27. Olson, J. A. Jr *et al.* American College of Surgeons Oncology Group Trials Z0010 and Z0011. Impact of immediate versus delayed axillary node dissection on surgical outcomes in breast cancer patients with positive sentinel nodes: results from American College of Surgeons Oncology Group Trials Z0010 and Z0011. *J. Clin. Oncol.* **26**, 3530–3535 (2008).
28. Caudle, A. S. *et al.* Multidisciplinary considerations in the implementation of the findings from the American College of Surgeons Oncology Group (ACOSOG) Z0011 study: a practice-changing trial. *Ann. Surg. Oncol.* **18**, 2407–2412 (2011).
29. Caudle, A. S. *et al.* American College of Surgeons Oncology Group (ACOSOG) Z0011: impact on surgeon practice patterns. *Ann. Surg. Oncol.* **19**, 3144–3151 (2012).
30. Alliance for Clinical Trials in Oncology. *Our vision* [online], https://www.allianceforclinicaltrialsinoncology.org (2015).
31. American College of Surgeons Oncology Group and Radiation Therapy Oncology Group: ACOSOG Z4099/ RTOG 1021. A randomized phase III study of sublobar resection (+/-brachytherapy) versus stereotactic body radiation therapy in high risk patients with stage I non-small cell lung cancer (NSCLC) http://atc.wustl.edu/protocols/rtog/1021/Z4099-1021_A0_05-02-2011 (2011)
32. Alliance for Clinical Trials in Oncology. *Alliance study result summaries* [online], https://www.allianceforclinicaltrialsinoncology.org/main/public/standard.xhtml?path=%2FPublic%2FResults (2015).
33. European Organisation for Research and Treatment of Cancer. *Clinical Trials Database* [online], http://www.eortc.org/clinical-trials/ (2015).
34. Jacquier, I., Boutron, I., Moher, D., Roy, C. & Ravaud, P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. *Ann. Surg.* **244**, 677–683 (2006).
35. National Institute for Health and Care Excellence. *NICE technology appraisal guidance [TA176]: Cetuximab for the first-line treatment of metastatic colorectal cancer* [online], https://www.nice.org.uk/guidance/ta176
36. Naredi, P. & La Quaglia, P. The future of trials in surgical oncology. *Nat. Rev. Clin. Oncol.* **12**, 425–431 (2015).
37. [No authors listed] The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Can. Med. Assoc. J.* **121**, 1193–1254 (1979).
38. Sackett, D. L. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* **95** (2 Suppl.), 2S–4S (1989).
39. Adam, R. *et al.* Patients with initially unresectable colorectal liver metastases: is there a possibility of cure? *J. Clin. Oncol.* **27**, 1829–1835 (2009).
40. Brouquet, A. *et al.* High survival rate after two-stage resection of advanced colorectal liver metastases: response-based selection and complete resection define outcome. *J. Clin. Oncol.* **29**, 1083–1090 (2011).
41. Elias, D. *et al.* Peritoneal colorectal carcinomatosis treated with surgery and perioperative intraperitoneal chemotherapy: retrospective analysis of 523 patients from a multicentric French study. *J. Clin. Oncol.* **28**, 63–68 (2010).
42. Curley, S. A. Radiofrequency ablation versus resection for resectable colorectal liver metastases: time for a randomized trial? *Ann. Surg. Oncol.* **15**, 11–13 (2008).
43. Barkun, J. S. *et al.* Evaluation and stages of surgical innovations. *Lancet* **374**, 1089–1096 (2009).
44. Köckerling, F. The need for registries in the early scientific evaluation of surgical innovations. *Front. Surg.* **1**, 12 (2014).
45. Schadde, E. *et al.* Early survival and safety of ALPPS: first report of the International ALPPS Registry. *Ann. Surg.* **260**, 829–836 (2014).
46. Boden, W. E., O'Rourke, R. A., Teo, K. K., Hartigan, P. M. & Maron, D. J. Optimal medical therapy with or without PCI for stable coronary disease. *N. Engl. J. Med.* **356**, 1503–1516 (2007).
47. Benson, K. & Hartz, A. J. A comparison of observational studies and randomized, controlled trials. *N. Engl. J. Med.* **342**, 1878–1886 (2000).
48. Concato, J., Shah, N. & Horwitz, R. I. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N. Engl. J. Med.* **342**, 1887–1892 (2000).
49. [No authors listed] Quality assurance in surgical oncology: the EURECCA polatform. *Eur. J. Surg. Oncol*. **40**, 1387–1390 (2014).
50. Moher, D. *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int. J. Surg.* **10**, 28–55 (2012).
51. Moher, D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *JAMA* **279**, 1489–1491 (1998).
52. Boutron, I., Moher, D., Altman, D. G., Schulz, K. F. & Ravaud, P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann. Intern. Med.* **148**, 295–309 (2008).
53. Nagendran, M. *et al.* Poor adherence of randomised trials in surgery to CONSORT guidelines for non-pharmacological treatments (NPT): a cross-sectional study. *BMJ Open* **3**, e003898 (2013).
54. Clavien, P. A. & Lillemoe, K. D. A new policy to implement CONSORT guidelines for surgical randomized controlled trials. *Ann. Surg.* **260**, 947–948 (2014).
55. Lacy, A. M. *et al.* Laparoscopy-assisted colectomy versus open colectomy for treatment of non-metastatic colon cancer: a randomised trial. *Lancet* **359**, 2224–2229 (2002).
56. Evrard, S., Mathoulin-Pelissier, S. & Kramar, A. Open versus laparoscopy-assisted colectomy. *Lancet* **361**, 73–76 (2003).
57. Buunen, M. *et al.* Survival after laparoscopic surgery versus open surgery for colon cancer: long-term outcome of a randomised clinical trial. *Lancet Oncol.* **10**, 44–52 (2009).
58. Kitchener, H., Swart, A. M., Qian, Q., Amos, C. & Parmar, M. K. Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study. *Lancet* **373**, 125–136 (2009).
59. Barton, D. P., Naik, R. & Herod, J. Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomized study. *Int. J. Gynecol. Cancer* **19**, 1465 (2009).
60. Nordlinger, B., Poston, G. J. & Goldberg, R. M. Should the results of the New EPOC trial change practice in the management of patients with resectable metastatic colorectal cancer confined to the liver? *J. Clin. Oncol.* **33**, 241–243 (2015).
61. Primrose, J. *et al.* Systemic chemotherapy with or without cetuximab in patients with resectable colorectal liver metastases: the New EPOC randomised controlled trial. *Lancet Oncol*. **15**, 601–611 (2014).
62. Evrard, S. *et al.* Unresectable colorectal cancer liver metastases treated by intraoperative radiofrequency ablation with or without resection. *Br. J. Surg.* **99**, 558–565 (2012).
63. McCulloch, P. *et al.* IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* **346**, f3012 (2013).
64. Simon R. Optimal two-stage designs for phase II clinical trials. *Control. Clin. Trials* **10**, 1–10 (1989).
65. US National Library of Medicine. *ClinicalTrials.gov* [online], https://clinicaltrials.gov/ct2/show/NCT01876901?term=Cascador&rank=1 (2015).
66. US National Library of Medicine. *ClinicalTrials.gov* [online], https://clinicaltrials.gov/ct2/show/NCT02218801 (2015).
67. [No authors listed] Improved survival with preoperative radiotherapy in resectable rectal cancer. Swedish Rectal Cancer Trial. *N. Engl. J. Med.* **336**, 980–987 (1997).
68. Peeters, K. C. *et al.* The TME trial after a median follow-up of 6 years: increased local control but no survival benefit in irradiated patients with resectable rectal carcinoma. *Ann. Surg.* **246**, 693–701 (2007).
69. den Dulk, M. *et al.* The abdominoperineal resection itself is associated with an adverse outcome: the European experience based on a pooled analysis of five European randomised clinical trials on rectal cancer. *Eur. J. Cancer* **45**, 1175–1183 (2009).
70. Nordlinger, B. *et al.* Perioperative chemotherapy with FOLFOX4 and surgery for resectable liver metastases from colorectal cancer. *Lancet* **371**, 1007–1016 (2008).
71. Kapiteijn, E. *et al.* Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N. Engl. J. Med.* **345**, 638–646 (2001).
72. van Gijn, W. *et al.* Volume and outcome in colorectal cancer surgery. *Eur. J. Surg. Oncol.* **36** (Suppl. 1), S55–S63 (2010).
73. den Dulk, M. *et al.* Improved overall survival for patients with rectal cancer since 1990: the effects of TME surgery and pre-operative radiotherapy.. *Eur. J. Cancer* **44**, 1710–1716 (2008).
74. Krijnen P, den Dulk, M., Meershoek-Klein-Kranenbarg, E., Jansen-Landheer, M. L. & van de Velde, C. J. Improved survival after resectable non-cardia gastric cancer in The Netherlands: the importance of surgical training and quality control. *Eur. J. Surg. Oncol.* **35**, 715–720 (2009).
75. West, N. P. *et al.* Pathology grading of colon cancer surgical resection and its association with survival: a retrospective observational study. *Lancet Oncol.* **9**, 857–865 (2008).
76. Quirke, P. *et al.* Effect of the plane of surgery achieved on local recurrence in patients with operable rectal cancer: a prospective study using data from the MRC CR07 and NCIC-CTG CO16 randomised clinical trial. *Lancet* **373**, 821–828 (2009).
77. Taylor, F. G. *et al.* Preoperative magnetic resonance imaging assessment of circumferential resection margin predicts disease-free survival and local recurrence: 5-year follow-up results of the MERCURY study. *J. Clin. Oncol.* **32**, 34–43 (2014).

Author contributions

S.E., P.M.-S., and G.P. researched the data for the article. All authors made substantial contributions to discussions of content. G.P. wrote the article and all authors reviewed/edited the manuscript before submission.