Review Article

Randomized controlled trials and alternative study designs in surgical oncology

Wim Ceelen [a, b, c], Kjetil Soreide [d, e, f, *]

[a] Department of GI Surgery, Ghent University Hospital, Ghent, Belgium
[b] Department of Human Structure and Repair, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium
[c] Cancer Research Institute Ghent (CRIG), Ghent, Belgium
[d] Department of Gastrointestinal Surgery, Stavanger University Hospital, Stavanger, Norway
[e] Department of Clinical Medicine, University of Bergen, Bergen, Norway
[f] SAFER Surgery, Surgical Research Unit, Stavanger University Hospital, Stavanger, Norway

ARTICLE INFO

ABSTRACT

Surgery is central to the cure of most solid cancers and an integral part of modern multimodal cancer management for early and advanced stage cancers. Decisions made by surgeons and multidisciplinary team members are based on best available knowledge for the defined clinical situation at hand. While surgery is both an art and a science, good decision-making requires data that are robust, valid, representative and, applicable to most if not all patients with a specific cancer. Such data largely comes from clinical observations and registries, and more preferably from trials conducted with the specific purpose of arriving at new answers. As part of the ESSO core curriculum development an increased focus has been put on the need to enhance research literacy among surgical candidates. As an expansion of the curriculum catalogue list and to enhance the educational value, we here present a set of principles and emerging concepts which applies to surgical oncologist for reading, understanding, planning and contributing to future surgeon-led cancer trials.

© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Surgery is central to the cure of most solid cancers and an integral part of modern multimodal cancer management for early and advanced stage cancers. Decisions made by surgeons and multidisciplinary team members are based on best available knowledge for the defined clinical situation at hand. While surgery is both an art and a science, good decision-making requires data that are robust, valid, representative and, applicable to most if not all patients with a specific cancer. Such data largely comes from observational studies and registries, and more preferably from trials conducted with the specific purpose of arriving at new answers.

Unfortunately clinical research driven by surgeons have been traditionally poor and even dubbed a 'comic opera' in the past [1]. Progress in surgical trials have been made over the past decades [2],

particularly being strong in Europe. However, the bulk of 'surgical science' is still made up of retrospective studies [3]. Also, the quality of surgical trials is still lacking several items to live up to the expected standards [4,5]. Historically, the quality of surgical RCTs has been rather low: a systematic review of 388 RCTs published between 2008 and 2020 showed that many of the included trials were small, studied minor clinical events, and were prone to several sources of bias [6]. There is an urgent need to remedy these issues.

Overall, trials are still few and far between in surgical research, also for cancer patients. Of note, a study from the United States using 2002 to 2008 California Cancer Registry data from more than 555.000 patients with stage I to IV solid organ tumors found that only 0.3% of the patient population with cancer were included in any type of clinical trials [7]. This very low number should be viewed in the context of the USA being the most prolific research country, with several countries lagging behind and with regions deprived of research altogether [8].

Surgical oncologists remain severely underrepresented in cancer research, particularly in high-impact journals [9]. The reasons for this lack of participation and collaboration in trials are well described (Fig. 1) and include limited funding, inadequate training,
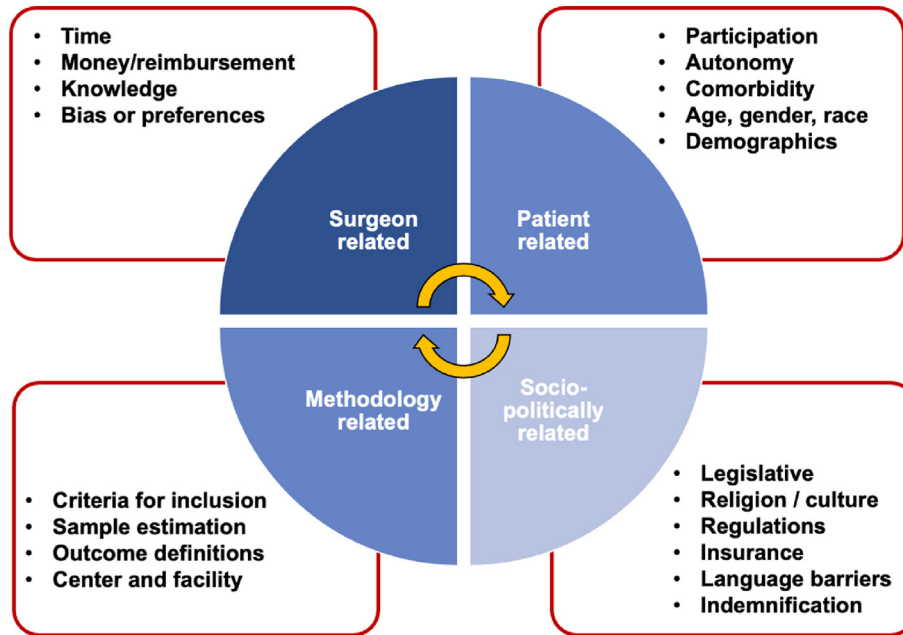
Fig. 1. Barriers to surgical trials and trials research.

lack of interest, lack of mentoring and support, and lack of time [10]. Nevertheless, a recent literature analysis shows that both the number of surgical clinical trials, and the proportion of 'late' trials (phase IV) are increasing [11]. In trials that investigate multi-modality regimens such as neoadjuvant or adjuvant therapy, surgical oncologists used to play a leading role. For example, the practice changing trials on neoadjuvant radiotherapy for locally advanced rectal cancer were surgeon driven. Unfortunately, surgical oncologists are much less involved nowadays in neoadjuvant trials, in particular when they are industry driven. It is imperative that the surgical oncology community retakes the initiative in this regard, and is involved as an equal partner in order to guarantee surgical quality and increase multi-disciplinary awareness.

As part of the ESSO core curriculum development [12], an increased focus has been put on the need to enhance research literacy among surgical candidates [13]. As an expansion of the curriculum catalogue list and to enhance the educational value, we here present a set of principles and emerging concepts which applies to surgical oncologist for reading, understanding, planning and contributing to future surgeon-led cancer trials.

2. The evidence hierarchy in medicine

Ever since the term "evidence-based medicine" entered the scientific lexography in the early 1990s [14], the strive to grade available data to support clinical care in a systematic way [15] has evolved through medical [16] and surgical practice [17]. Study designs are usually structured in a evidence hierarchy (Fig. 2), with the randomised controlled trial (RCT) as the gold standard for determining a difference between two treatments.

However, many research questions cannot be answered through an RCT [18]. Hence, the appreciation for observational studies, epidemiological research, frameworks for evaluating novel devices and surgical procedures [19,20], and mechanistic research (e.g from lab-driven data) are equally important for the progress of medicine.

Here, we will focus on the RCT and some of the emerging techniques and alternatives that are available to surgical oncologists to pursue randomized and quasi-randomized trial design methods [21].
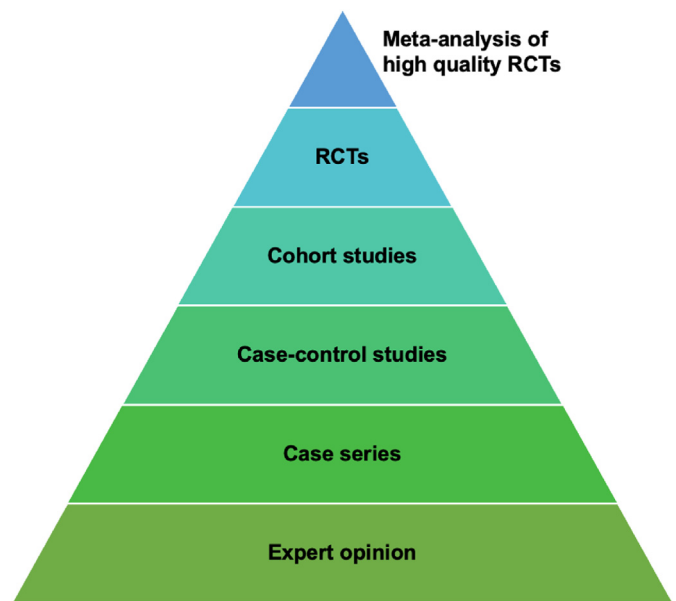

Fig. 2. The evidence-based medicine hierarchy pyramid.

3. Research question and trial design

The design of a RCT should guarantee that the results have internal validity and, avoid as much as possible sources of systematic error known as bias (Table 1). The first step in any scientific experiment is to clearly define the research question. In clinical studies, the research question is usually formatted as a PICO (definition of the patient population, interventional treatment, control treatment, and outcome). The research hypothesis should be Feasible, Interesting, Novel, Ethical, and Relevant ('FINER').

Next, the researcher should decide whether a superiority, equivalence, or non-inferiority design is appropriate. Typically, the non-inferiority RCT design is chosen when the experimental

Table 1
Sources of bias in randomized trials and how to avoid them.

Potential source of bias	Solution
HARK-ing; Hypothesizing After the Results are Known'	Study should be registered **before** trial start
Selection bias	Adequate random allocation; stratification
Performance bias	- Masking/blinding of patients and surgeons
Pygmalion effect	- Quality assurance
Attrition bias	Intention to treat analysis
Gender bias	Pre-plan analysis of gender effects; stratification
Detection bias	Masking/blinding of outcome assessors
Inflation of type I error rate	No pre-planned subgroup analyses
Risk of futility or excess toxicity	Plan an interim analysis

method is less invasive or more cost-effective compared to an accepted gold standard but may put patients at risk. An example are the trials that have compared minimally invasive with open surgery for colon cancer [22]. The experimental treatment is declared non-inferior when the 95% confidence interval around the effect estimate does not cross a predefined margin [23]. The outcome difference and concepts compared to superiority trials is shown in Fig. 3. Several surgical trials have been conducted with this methodology [24–27], which has its own caveats and limitations to consider [28].

3.1. Pragmatic versus explanatory trials

Despite their theoretical methodological superiority, RCTs inform only a fraction of daily clinical practice. The astronomical costs and complexity of RCTs have prompted the development of pragmatic trials, which offer an easier and cheaper design to generate evidence. Also, it is well known that RCTs often lack external validity, and trial results are often not replicated in the wider patient population. While explanatory trials primarily focus on a biological mechanism, pragmatic trials focus on external validity, and aim to show effectiveness in 'real world' settings rather than efficacy. Pragmatic trials utilize less strict inclusion criteria compared to 'explanatory' trials. One of their key attributes is that they aim to inform stakeholders (patients, clinicians, and decision makers) rather than elucidate a biological or pathophysiological mechanism. The extent of pragmatism of a trial can be quantified using the PRagmatic-Explanatory Continuum Indicator Summary 2 (PRECIS-2) score, which ranges from 9 to 45 (with a higher score reflecting increased pragmatism) [29]. A limitation of pragmatic trials is that they are inherently more prone to heterogeneity, and this must be carefully considered and accommodated in the planning, conduct and analysis of a pragmatic trial. Conversely, an explanatory trial may answer a very specific question by a stringent hypothesis yet have very low validity outside the narrow criteria for inclusion in the given trial. Hence, real-world applications from a purely explanatory trial result may be limited or the effect diluted when introduced to a wider population without validation. The true design of many trials, may not be purely explanatory nor purely pragmatic, as demonstrated across the field of critical care [30]. Also, problems with reporting seems to be similar in both pragmatic and explanatory trials in surgery [31]. Some key issues to consider when designing surgical interventions include the identification of each surgical intervention and their components, who
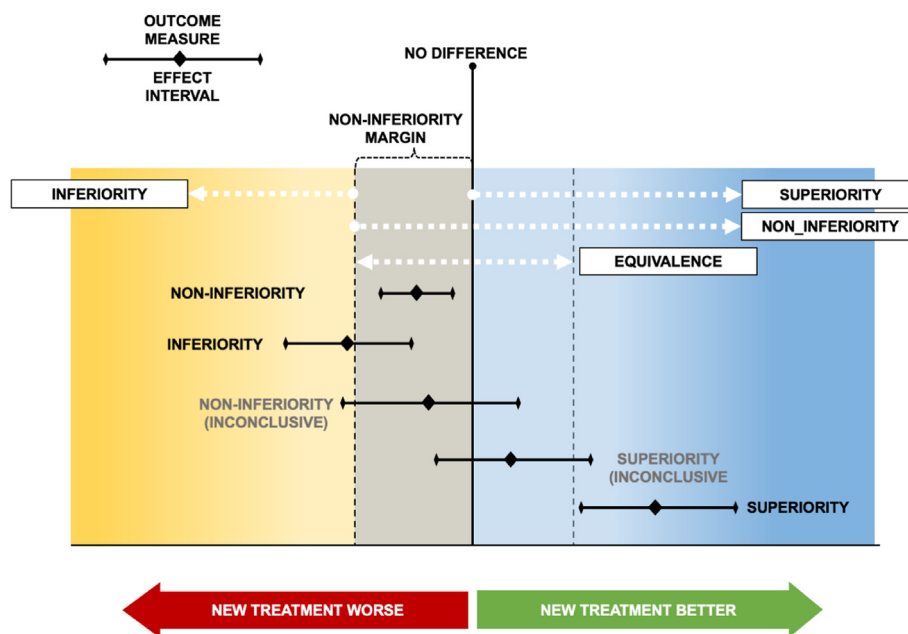
Fig. 3. Point estimate interval for different outcomes using superiority, equivalence and non-inferiority methods.
Legend:
The different effect measures and their direction that are used for different trial designs, whereby superiority design implies an improvement of new over previous or standard treatment. Non-inferior design relies on the effect being within the defined non-inferior margin. Equivalence trials are designed to show equal effect within a defined margin. Point estimates with bars represent tentative situations and interpretation of the results.

will deliver the interventions, and where and how the interventions will be standardized and monitored during the trial. The trial design (pragmatic and explanatory), comparator and stage of innovation may also influence the extent of detail required [32]. In order for evidence from trials to be used in clinical practice and policy, trialists should make every effort to make trials widely applicable, which means that more trials should be designed towards a pragmatic design in attitude [33]. Of note, even where surgical RCTs exist, these seem to have variable impact on adoption to clinical practice [34].

4. Choice of study endpoints

4.1. Primary endpoint

The choice of a relevant primary endpoint is one of the most important steps in the design of a RCT. The sample size and power calculation of a randomized trial are based on the estimated effect of the intervention on this primary endpoint. All too often, surgical trials use endpoints chosen for convenience rather than patient relevance. Examples include the volume of drainage fluid, or the duration of a procedure.

In oncology trials, it is common to use *surrogate endpoints*, which are intermediate on the pathway between the exposure or intervention and the 'hard' endpoint, such as death or long-term survival. An example is recurrence-free survival (RFS) often reported rather than overall survival (OS), demonstrated recently in surgery for colorectal liver metastasis to be largely different and RFS to be an invalid endpoint compared to OS [35].

The obvious advantage of using surrogate endpoints is that more (and earlier) events will occur in the same time-period, thus enhancing the power of the study and/or allowing to limit its duration and costs. However, it should be kept in mind that despite the increasing use of these surrogates, they are often unvalidated and do not adequately predict the 'hard' endpoint [35,36]. Further examples of surrogate endpoints that are unvalidated include the rate of pathological response after neoadjuvant therapy in rectal cancer (rather than survival), and the harvested lymph node count in colon cancer (as a proxy for 'quality of surgery' rather than true survival) [37–39].

4.2. Secondary endpoints

Examples of secondary endpoints include response data, surgical outcome, or laboratory data. Typically, observed effects on secondary outcomes should be interpreted cautiously because of the inflated type I error rate associated with multiple testing.

4.3. Patient reported outcome measures

Increasingly, funders require that due attention is given to involvement of patients or patient organizations in the design, execution, and reporting of trials. As a minimum, prospective clinical trials should include patient reported outcome measures (PROMs). Typically, this entails the use of validated questionnaires that prospectively follow patient's perceived health related quality of life.

4.4. Composite endpoints

Some endpoints or outcomes are too rare to meaningfully investigate alone. One suggested solution is to generate a 'composite' endpoint, which is considered to have occurred when one of the component events occurs. One such composite endpoint was developed for liver surgery [40]. The components selected for the liver surgery-specific composite endpoint were ascites, liver failure, bile leakage, intra-abdominal haemorrhage, intra-abdominal abscess, and operative mortality, all with a Clavien-Dindo grade of at least 3 and occurring within 90 days after initial surgery. The composite endpoint is reached when any of these events occurs. A disadvantage of using composite endpoints is the fact that it assigns equal importance to endpoints with a very different relevance for patients, e.g. a postop bile leak or mortality in the example given.

4.5. Core Outcome Measures in Effectiveness Trials (COMET) initiative

In many fields, considerable heterogeneity exists in outcome measurement among trials which study the same intervention. This hampers the ability to compare and pool results from individual studies. The COMET initiative (Core Outcome Measures in Effectiveness Trials, www.comet-initiative.org) was initiated in 2010 in order to stimulate the development and application of agreed standardised sets of outcomes (core outcome sets) for a specific disease or treatment.

5. Sample size calculation

Calculation of the required sample size to demonstrate a causal effect of the intervention is an essential step in the design of a RCT. Calculation of the sample size requires the following information: the study design, the estimated effect of the experimental intervention (e.g. we hypothesize that the novel intervention will improve the 5 year survival with 10% on average, compared to controls), the precision (variance) of this estimate, the type I error rate (typically 5%), and the desired power (typically 80%). The effect size estimate and its variance should be carefully chosen, and can be based on previous studies, literature data, preclinical data, or a pilot experiment [41]. The DELTA^2 guideline provides guidance on determining the target difference and sample size calculation in RCTs, as well as recommendations for reporting the sample size calculation [42]. Many surgical studies have a limited sample size, and can therefore detect only large effects. Underpowered studies will not only fail to detect a potentially relevant effect, but will also overestimate the effect size, if an effect is observed. Data suggest that these issues are not adequately addressed in many surgical trials [43].

6. Randomization and masking

Studies using random treatment allocation are commonly regarded as being able to generate the highest level of certainty regarding the effects of an intervention (Fig. 2). The goal of randomization is to control for known as well as unknown potential confounders, ensuring that any observed treatment effect is the result of the intervention. Although it is often stated that the aim of randomization is to create balanced groups, its main role is to allow valid estimation of standard errors [44]. Some remaining imbalances are unavoidable, but these are by definition the result of chance. In order to guarantee treatment allocation masking, patients should be randomized at a timepoint as close as possible to the actual intervention. Several 'pseudo' randomization methods such as treatment allocation according to the day of the week, date of birth, etc violate the principle of allocation blinding. Acceptable methods are software generated random lists or the use of opaque sealed envelopes.

Specifically for pragmatic or small trials, it should be considered to *stratify* randomization. Patients are then randomized within strata that represent one or more factors known to be associated with the outcome. As an example, pragmatic multicenter trials

often stratify randomization on centre, since between centre differences may create a baseline imbalance.

Failure to mask researchers and patients to the allocated intervention leads to multiple sources of bias and overestimation of the treatment effect [45]. Specifically, masking (blinding) of surgeons and patients is essential to avoid performance bias and the Pygmalion effect. The latter refers to the fact that the expectation of the surgeon, when she knows the treatment that will be assigned, may affect patient outcome, thus creating a bias [46]. While masking of surgeons to surgical procedures is evidently impossible, with some efforts patients have been masked to open versus laparoscopic procedures by covering the entire abdomen with a bandage. When masking of neither surgeons nor patients is feasible, those assessing the outcomes should be masked.

Trials that compare a surgical with a medical treatment are confronted with specific challenges in recruitment and retention of patients. Both the surgeon and the patient may have strong *a priori* preferences, and perceive a lack of equipoise, which makes trial participation challenging (Fig. 4) [47].

A specific form of clinical trial evaluating a surgical technique are those where the control arm consists of 'sham' surgeries or placebo interventions. A historical example is a trial that studied the effect of ligation of the internal mammary artery to treat angina pectoris [48]. In both groups, the internal mammary arteries were dissected and looped, but only in the experimental group they were ligated. No differences in outcome were observed. A more recent example is a RCT evaluating arthroscopic subacromial decompression for subacromial shoulder pain, in which patients were assigned to either arthroscopic subacromial decompression, arthroscopy only (sham surgery group), or no treatment [49]. Interestingly, the outcome in the decompression arm was equivalent to that in the sham surgery arm. Obviously, sham controlled

trials are not relevant for cancer-related questions where the intervention is intended to cure.

7. Analysis and reporting

### 7.1. Intention to treat versus per protocol analysis

The aim of the 'intention to treat' (ITT) approach is to avoid attrition bias caused by an uneven rate of dropouts between study arms. In the ITT analysis, outcomes are analyzed according to the treatment allocated, irrespective of whether a patient actually received the treatment. When a considerable amount of missing data or protocol deviations are present, authors increasingly deviate from the ITT approach ('modified' ITT analysis). However, this practice was shown to result in overestimation of the treatment effect, and multiple imputation to address missing values is the preferred approach [50].

7.2. Sex and gender effects

It is increasingly realized that both sex and gender can interact with the treatment effect in clinical trials [51]. Therefore, funders such as the European Commission require that due attention is paid to the analysis of the results according to sex and gender [52].

7.3. Reporting of 'time to event' outcomes

Oncology trials often study a time dependent outcome such as overall or progression free survival. Typically, some patients have not experienced the event of interest by the time of analysis or are lost to follow up and, are therefore right-censored. The survival probability can be estimated using the Kaplan-Meier product-limit
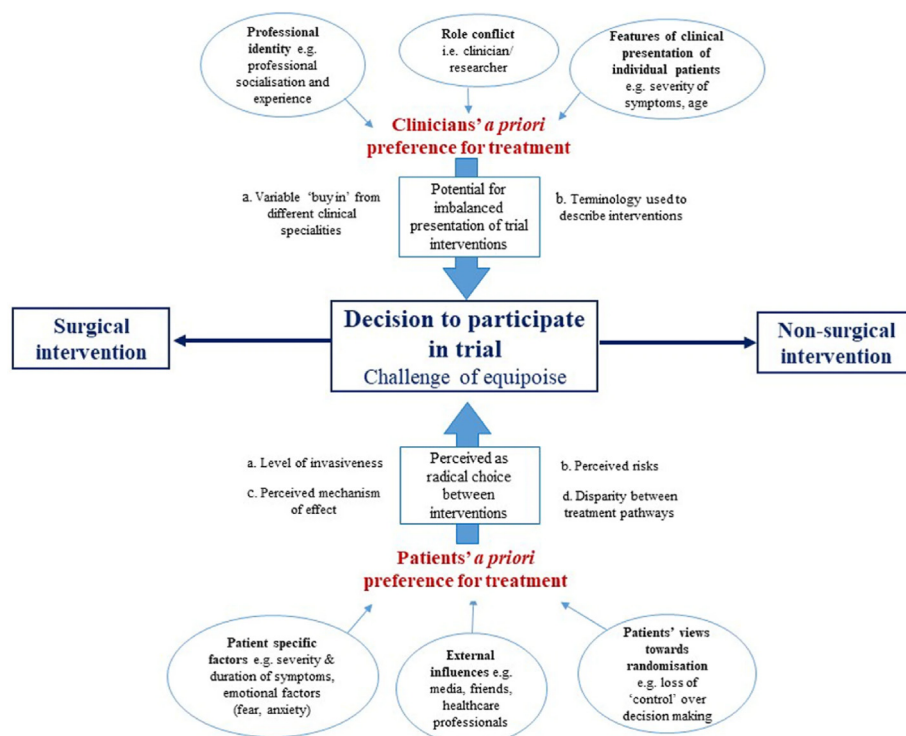
Fig. 4. Challenges associated with randomized trials comparing a surgical with a non-surgical intervention.
Legend: A priori preferences among clinicians and patients in trials comparing surgery to non-surgical intervention.
Reproduced under the Creative Commons licence ©Authors from Davies, L., Beard, D., Cook, J.A. et al. The challenge of equipoise in trials with a surgical and non-surgical comparison: a qualitative synthesis using meta-ethnography. *Trials* **22**, 678 (2021).

method [53], which does not make any assumptions about the underlying shape (e.g. exponential decline) of the survival curve (it is 'nonparametric'). When reporting a Kaplan-Meier estimate, the numbers at risk should be provided along the X-axis, and the survival curve should ideally be supplemented with its associated 95% confidence interval.

Comparisons between survival curves are usually tested with the log-rank test, which tests the hypothesis that there is no difference between groups in the probability of the event at any time. Importantly, similar to the Kaplan-Meier estimator, the log-rank test assumes that censoring is 'non informative', i.e. that censored subjects have the same probability of experiencing the event. In surgical studies, this assumption can be violated when patients drop out of the study (and are censored) due to excessive complications, which obviously puts these patients at higher risk of experiencing the event of interest. Importantly, the logrank test loses its ability to detect differences between survival curves if the hazards between both groups are *non-proportional* over time, which becomes evident when survival curves are *crossing*. Such a scenario has become more frequent in studies of immune checkpoint inhibitors, which often show an initially poorer survival, but later a better survival compared to standard therapy [54]. When crossing survival curves are encountered, specialist statistical techniques and guidance are required [55]. The Kaplan-Meier test provides a *P*-value, but not the *size* of the difference between groups in time to event outcome. The effect size can be estimated using the Cox model, which calculates the ratio of the hazard in both treatment arms. The hazard function is the *instantaneous event rate*, and mathematically the derivative of the survival function.

7.4. Reporting of trials

The reporting of randomized trials should adhere to the CONSORT (Consolidated Standards of Reporting Trials) guidelines (www.consort-statement.org). In 2008, a CONSORT extension for nonpharmacologic treatments (including surgery) was published and further updated in 2017 [10]. Depending on the trial design, many other extensions have been published, and these can be found at the website of the Equator network (www.equator-network.org).

7.5. Registration of trials

In the past, it was found that half of all published trials had at least one primary outcome changed, newly introduced, or omitted compared to protocol [56]. This lead to a form of scientific misconduct known as HARKing, or 'Hypothesizing After the Results are Known', in order to make the data fit the desired outcome [57]. Also, negative results often remain unpublished, leading to publication or reporting bias. To try to prevent both sources of bias, journals and funders now require that trial protocols are registered *before the start of the study* in a national or international public repository such as www.clinicaltrials.gov.

8. Trials evaluating devices, implants, or surgical procedures

Before the implementation of the novel Medical Device Regulation (MDR) in the EU, very few novel devices, implants, or surgical procedures were evaluated in formal clinical trials, leading to high profile scandals such as those involving the PIP silicone breast prosthesis, vaginal meshes, and faulty hip prostheses [58,59]. The absence of formal evaluation in clinical studies can be explained by Buxton's law: 'it is aways too early, until it is suddenly too late' [60]. This refers to the fact that it is often felt that formal evaluation of a novel device is not yet desirable because of learning

curves and ongoing technical improvements, until at a certain point the pressures of personal prestige and the industry led to widespread clinical adoption without proof of safety. The MDR now makes clinical studies compulsory for all high-risk (class III and class II implantable) devices.

The Idea, Development, Exploration, Assessment, and Long-term Follow-up (IDEAL) framework and recommendations were introduced as a structured approach that allows to reconcile surgical innovation and protection of the patient [61]. Originally designed to evaluate surgical techniques, it was later extended to the study of devices (IDEAL-D). It has been argued that devices which do not claim superior efficacy and do not claim to be new in terms of mechanism of action need not be subjected to randomized trials [18].

9. Alternative trial designs in surgical oncology

While randomized trials allow unbiased estimation of average treatment effects and are regarded as the design offering the highest internal validity, their results often suffer from poor *external* validity [33,62,63]. Specifically, it is known that the characteristics of the highly selected patients participating in clinical trials usually differ from the average patient, resulting in the observation that results from RCTs are often not replicated in the 'real world' setting.

Also, surgeons and patients may find it difficult to accept random allocation of treatment when there is no perceived equipoise between the intervention and the control treatment. This typically arises when a surgical intervention is compared with a medical treatment, or when open and minimally invasive surgery are compared (Table 2). Another problem is the issue with poor outcomes in the control arm [64], e.g. suggesting that the standard of care has not been met for controls over the intervention arm. Notably, several modified or alternative trial designs have been proposed to overcome the obstacles associated with perceived lack of equipoise [21].

9.1. Cluster randomized trials

In cluster randomized trials, the unit randomized is not an individual subject, but a group (cluster) of subjects such as a hospital, ward, or geographical entity. They are mainly used to study public health interventions such as guideline implementation or quality improvement initiatives. Also, they may be the preferred design when there is a high risk of bias by *contamination*. As an example, when a surgical oncologist is recruiting patients for an RCT comparing open versus minimally invasive surgery, overhearing other patients in the waiting room may influence the willingness of potential trial candidates to participate in the trial. In this setting, randomizing hospitals rather than patients, may sequentially allow for a change in practice yet allow for measuring the effect from the 'old' to the 'new' strategy.

9.2. Stepped-wedge trial-design

A specific cluster randomized trial design is the stepped wedge trial (Fig. 5), which gradually introduces the novel intervention to groups of clusters over time [21,65]. Groups of clusters cross over to the experimental intervention in random order and in a stepwise fashion, until all groups have received the experimental intervention at the end of the trial. An example is a study introducing the surgical safety checklist which was sequentially rolled out in a random order until all 5 clusters in two hospitals had received the checklist; the results showed a significant reduction in morbidity and length of stay [66]. Another example is a trial aiming at

Table 2
Key differences between oncology trials investigating an investigational medicinal product (IMP) versus a surgical technique or device.

Key differences	Trials investigating an investigational medicinal product (IMP)	Trial investigating a surgical technique, device, or implant
Investigational product	Pharmaceutical product	Surgical technique, device, or implant
Compliance	Depends on patient, remains a concern throughout the trial	Depends on surgeon, relevant during the index surgery
Masking	Usually possible with matching placebo	Usually not possible, but assessors may be masked
Cointerventions	A risk in drug studies, may be inadvertent	Less of a risk, may be intentional
Crossover	Uncommon, may occur inadvertently through packaging errors	A common concern, may indicate surgeon expertise or individual equipoise
Early-phase studies	Phase 1, phase 2, regulated by health agencies	Case series, unregulated by health agencies
Interventions	Very reproducible, biological activity may depend on genetic diversity	Intervention not as precisely defined, may vary depending on surgical expertise
Generalizability	Recruitment of population with minimal exclusion criteria from various types of centers	Inclusion of a variety of surgeons, and loosely defined interventions
Feasibility	Depends on numbers of patients with condition	Can also depend on numbers of surgeons with expertise with new procedure

increasing patients' engagement in breast cancer surgery decision-making through a shared decision-making intervention [67]. Reported stepped wedge trials were found to often lack methodological quality, including failure to incorporate time effects and repeated measures in power calculations [68].

9.3. Patient-preference trials

One of the main obstacles confronting surgical trials is the perceived lack of equipoise between treatment arms. This has motivated the Zelen design [69], characterized by randomization *prior to consent,* followed by encouragement to accept the assigned treatment. Randomizing patients prior to consent is potentially problematic from an ethical point of view, and therefore this design is usually implemented to evaluate the benefit of *adding* a potentially beneficial treatment or service to usual care. In surgical oncology, an example is a randomized trial comparing radiofrequency ablation with surgical resection in the treatment of hepatocellular carcinoma, where the authors used the Zelen design to avoid patient refusal [70].

In the partially randomized patient preference design (PRPD), patients are asked whether they have a treatment preference. [71] If they do not, they are randomized; if they do have a clear treatment preference, they are offered the treatment of choice. A recent example is the protocol of the STAR-TREC trial, which will offer patients the choice between TME and organ preservation for early stage rectal cancer [72]. Those who prefer organ preservation will be allocated to either long course chemoradiation versus short-course radiotherapy, with selective transanal microsurgery. With the recent emphasis on patient participation, this trial design is increasingly used, and has shown to result in comparable outcomes between the randomized cohort and the preference cohort [71].

9.4. Expertise-based trials

In certain circumstances, surgeons may be reluctant to participate in a trial which randomizes patients to two different surgical techniques, if she is an expert in one of these. Typically, this setting occurs in comparative trials with minimal-invasive to open surgery techniques. Also, surgeons may be in the learning curve of the procedure. These restrictions may be overcome by the expertise-based trial design, where surgeons only provide the procedure in which they have appropriate expertise [73]. Patients therefore receive treatment from different surgical teams or hospitals, according to the allocated treatment. A recent example is the MIVATE trial, which will evaluate minimally invasive versus open esophagectomy for esophageal carcinoma [74]. The allocated procedure will be performed by the most experienced surgeon in each surgical center.

9.5. Registry-based trials

In registry based pragmatic trials [75], data collection and analysis uses existing large scale registries [76], which significantly simplifies the administrative and financial burden of a large RCT [75—79]. A prerequisite is the presence of a prospective registry, run either nationally, regionally or through a society [75,77,79,80]. An example is the TASTE trial, which assigned patients with myocardial infarction to either conventional percutaneous coronary intervention (PCI) or to thrombus aspiration followed by PCI, and used the existing Swedish angiography and angioplasty registry (SCAAR) as a platform [81]. The PyloResPres registry based trial will compare pylorus preservation versus pylorus resection in patients undergoing pancreaticoduodenectomy, and will use the StuDoQ| Pancreas registry established by the German Society of General and Visceral Surgery as a trial platform [82]. Registry-based RCT design can also be helpful for rare or orphan cancers, such as adrenocortical carcinoma. The European Network for the Study of Adrenal Tumors (ENSAT) has described three different registry-based RCTs through which they aim to enhance the evidence for treating this rare tumor [83].

9.6. Trials within cohorts

In this design, broad informed consent for data collection is obtained before enrollment in a prospective cohort. A second informed consent may be asked when a patient is randomized to the intervention arm; if they are assigned to the standard arm patients are not informed. As an example, the MEDOCC-CrEATE randomized trial will be performed within the Prospective Dutch ColoRectal Cancer cohort (PLCRC) [84]. The subgroup of stage II patients will be randomized to either standard follow up or an experimental arm, consisting of chemotherapy when circulating tumor DNA is detected.

9.7. Adaptive trial designs

The protocol of the standard RCT is fixed, and outcomes or methods must not be changed during the trial course. In contrast, an adaptive trial entail continuous modifications to key components of trial design (allocation ratio, sample size, eligibility criteria, number of treatment arms) while data are being collected [85]. Advantages of the adaptive design include reduced cost, shorter time to trial completion, lower risk of allocating patients to an
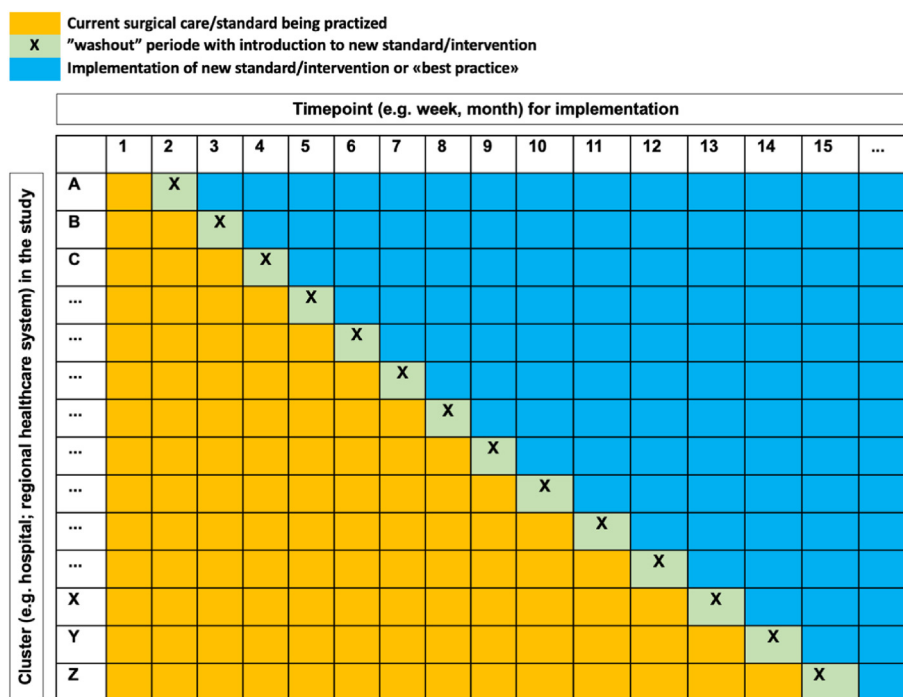
Fig. 5. Stepped-wedge design outline.
Legend: each unit of hospital (letters A to Z) are recruiting patients in standard care (yellow zone) over time (i.e. weeks, months, marked by numbers 1, 2, 3, …etc), then randomized to intervention (blue zone) which is implemented during a "washout" period (green). Eventually the trial will have recruited patients in 'standard' and 'intervention' groups for comparison.

ineffective treatment, and greater scientific and clinical relevance of the results. Typically, they use Bayesian methods to continuously adapt design features based on the observed outcomes. Adaptive designs have mostly been used in drug development, but some examples are available in the field of surgery, e.g. a Bayesian adaptive clinical trial of tranexamic acid in severely injured children [86]. Related is the group sequential design, which does not calculate a sample size but evaluates data as they are collected; the trial is stopped when a predefined threshold of efficacy, futility, or toxicity is reached [87].

10. Novel developments in surgical oncology trials

10.1. Precision oncology in surgery

Cancer therapy is increasingly guided by genomic and molecular biomarkers. This has led to the development of novel biomarker guided trial designs such as *basket trials*, which include patients with different cancer types but sharing a common molecular alteration, and *umbrella trials*, in which multiple targeted therapies are tested in a single cancer type according to the molecular profile identified [88]. It is likely, that the indications for surgery, the extent or type of surgery, and the use of neoadjuvant and adjuvant therapy will also be increasingly based on molecular biomarkers, and not only on imaging and routine clinicopathological staging. As an example, nomograms based on expression of S100A2 or S100A4 allow to identify pancreatic cancer patients with a particularly poor prognosis, and to avoid futile surgery [89]. Another example is the use of targeted imaging probes, based on a tumor's molecular profile, to guide the surgical approach or to define margin status intraoperatively [90]. Furthermore, the administration of neoadjuvant therapy presents new challenges, as tumors may show a complete or near-complete response. In this setting, there may be a

need to randomize for further strategy: should one proceed with surgery as planned or, enroll patients in a watch and wait type of surveillance program? A more personalized strategy with emergence of new trial designs may allow to identify the best method to the specific question for such clinical challenges.

10.2. Artificial intelligence (AI) driven cancer trials

The increasing availability of multimodal 'omics' data sets and the development of computing power and bioinformatic infrastructure has resulted in the development of in silico trials, which use a variety of artificial intelligence tools to emulate cancer trials. While these in silico methods cannot replace clinical trials, they allow to inform the design of a study, resulting in improved efficacy and safety [91].

Furthermore, the interest into radiomics and deeper understanding of data derived from CT, MRI, PET scans and other imaging studies may yield prognostic signatures that may be used to stratify into different treatment arms in the future, also for surgical patients. Hence, several emerging technologies may influence surgical options, strategies and methodology for trial execution.

11. Conclusions

In modern cancer care, the surgical oncologist is at the centre of several ground-breaking discoveries and emerging technologies. While the standard RCT design remains the reference standard for clinical experimentation, alternatives are on the horizon, but implementation in surgical clinical research is still embryonic. An increased recognition by surgeon-scientists, surgeon-trialists and among surgical journal editors of these alternative prospective trial designs and of their methodological requirements are imperative. Only by investing in research methods we can hope to recruit more

patients to appropriately designed trials and hence arrive at new answers for more optimal and personalized surgical cancer treatment.

Funding

CRediT authorship contribution statement

Wim Ceelen: Conceptualization, Methodology, Data curation, Writing — original draft, Writing — review & editing, Final approval for submission. **Kjetil Soreide:** Conceptualization, Methodology, Data curation, Figures, Visualization, Writing — review & editing, Final approval for submission.

Declaration of competing interest

None reported.

References

[1] Horton R. Surgical research or comic opera: questions, but few answers. Lancet 1996;347:984—5.

[2] Ahmed Ali U, et al. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. Ann Surg 2013;258:199—207.

[3] Rosenberg J. Surgical research remains comic opera and maybe for good reasons. Dan Med J 2021;68.

[4] Ahmed Ali U, et al. Journal impact factor and methodological quality of surgical randomized controlled trials: an empirical study. Langenbeck's Arch Surg 2017;402:1015—22.

[5] Papet E, et al. Methodological and ethical quality of surgical trials from 2016 to 2020. Langenbecks Arch Surg; 2022.

[6] Robinson NB, et al. Characteristics of randomized clinical trials in surgery from 2008 to 2020: a systematic review. JAMA Netw Open 2021;4:e2114494.

[7] Chow CJ, et al. Does enrollment in cancer trials improve survival? J Am Coll Surg 2013;216:774—80. discussion 80-1.

[8] Are C, et al. Global variations in the level of cancer-related research activity and correlation to cancer-specific mortality: proposal for a global curriculum. Ann Surg Oncol 2018;25:594—603.

[9] Holliday EB, et al. Does cancer literature reflect multidisciplinary practice? A systematic review of oncology studies in the medical literature over a 20-year period. Int J Radiat Oncol Biol Phys 2015;92:721—31.

[10] Naredi P, La Quaglia MP. The future of trials in surgical oncology. Nat Rev Clin Oncol 2015;12:425—31.

[11] Niforatos JD, et al. Trends in published clinical trial phases among surgical specialties, 2007-2017. Ann Surg 2021;273:e178—9.

[12] van der Hage J, et al. The ESSO core curriculum committee update on surgical oncology. Eur J Surg Oncol 2021;47:e1—30.

[13] Are C, et al. Global curriculum in research literacy for the surgical oncologist. Eur J Surg Oncol 2018;44:31—42.

[14] Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992;268:2420—5.

[15] Guyatt GH, et al. Users' guides to the medical literature: IX. A method for grading health care recommendations. JAMA 1995;274:1800—4.

[16] Montori VM, Guyatt GH. Progress in evidence-based medicine. JAMA 2008;300:1814—6.

[17] Meakins JL. Evidence-based surgery. Surg Clin 2006;86:1—16 [vii].

[18] Páez A, et al. Beyond the RCT: when are randomized trials unnecessary for new therapeutic devices, and what should we do instead? Ann Surg 2022;275:324—31.

[19] Dimick JB, et al. The IDEAL framework for evaluating surgical innovation: how it can Be used to improve the quality of evidence. JAMA Surg 2019;154:685—6.

[20] Fleetcroft C, et al. IDEAL as a guide to designing clinical device studies consistent with the new European Medical Device Regulation. BMJ Surg Interv Health Technol 2021;3:e000066.

[21] Augustinus S, et al. Alternative randomized trial designs in surgery: a systematic review. Ann Surg 2022;276:753—60.

[22] Buunen M, et al. Survival after laparoscopic surgery versus open surgery for colon cancer: long-term outcome of a randomised clinical trial. Lancet Oncol 2009;10:44—52.

[23] Cuzick J, Sasieni P. Interpreting the results of noninferiority trials-a review. Br J Cancer 2022 Nov;127(10):1755—9.

[24] Shu C, et al. Surgical clinical trials with non-inferiority design: a cross-sectional bibliometric analysis. Ann Transl Med 2021;9:1302.

[25] Carneiro VCG, et al. Proof-of-concept randomized phase II non-inferiority trial of simple versus type B2 hysterectomy in early-stage cervical cancer ≤2 cm (LESSER). Int J Gynecol Cancer 2023. online ahead of print.

[26] He L, et al. Surgical site infection after intracorporeal anastomosis for left-sided colon cancer: study protocol for a non-inferiority multicenter randomized controlled trial (STARS). Trials 2022;23:954.

[27] Nilsson M, et al. Neoadjuvant chemoradiotherapy and surgery for esophageal squamous cell carcinoma versus definitive chemoradiotherapy with salvage surgery as needed: the study protocol for the randomized controlled NEEDS trial. Front Oncol 2022;12:917961.

[28] Ranstam J, Cook JA. Non-inferiority and equivalence trials. Br J Surg 2017;104:1578—9.

[29] Loudon K, et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147.

[30] Palakshappa JA, et al. Systematic review of the "pragmatism" of pragmatic critical care trials. Crit Care Explor 2022;4:e0738.

[31] Blencowe NS, et al. Systematic review of intervention design and delivery in pragmatic and explanatory surgical randomized clinical trials. Br J Surg 2015;102:1037—47.

[32] Blencowe NS, et al. Interventions in randomised controlled trials in surgery: issues to consider during trial design. Trials 2015;16:392.

[33] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials 2009;10:37.

[34] Oberkofler CE, et al. Procedural surgical RCTs in daily practice: do surgeons adopt or is it just a waste of time? Ann Surg 2019;270:727—34.

[35] Ecker BL, et al. Recurrence-free survival versus overall survival as a primary endpoint for studies of resected colorectal liver metastasis: a retrospective study and meta-analysis. Lancet Oncol 2022;23:1332—42.

[36] Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Med 2017;15:134.

[37] Petrelli F, et al. Pathologic complete response and disease-free survival are not surrogate endpoints for 5-year survival in rectal cancer: an analysis of 22 randomized trials. J Gastrointest Oncol 2017;8:39—48.

[38] Mathis KL, et al. Surgical quality surrogates do not predict colon cancer survival in the setting of technical credentialing A report from the prospective cost trial. Ann Surg 2013;257:102—7.

[39] Veen T, et al. Qualitative and quantitative issues of lymph nodes as prognostic factor in colon cancer. Dig Surg 2013;30:1—11.

[40] van den Broek MA, et al. Development of a composite endpoint for randomized controlled trials in liver surgery. Br J Surg 2011;98:1138—45.

[41] Fairhurst K, et al. Value of surgical pilot and feasibility study protocols. Br J Surg 2019;106:968—78.

[42] Cook JA, et al. DELTA(2) guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. BMJ 2018;363:k3750.

[43] Yu J, et al. Design, conduct, and analysis of surgical randomized controlled trials: a cross-sectional survey. Ann Surg 2019;270:1065—9.

[44] Senn S. Seven myths of randomisation in clinical trials. Stat Med 2013;32:1439—50.

[45] Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002;359:614—8.

[46] Mossad EB. To every nerve there is a needle-the threat of a pygmalion effect. J Cardiothorac Vasc Anesth 2021;35:137—8.

[47] Davies L, et al. The challenge of equipoise in trials with a surgical and non-surgical comparison: a qualitative synthesis using meta-ethnography. Trials 2021;22:678.

[48] Cobb LA, et al. An evaluation of internal-mammary-artery ligation by a double-blind technic. N Engl J Med 1959;260:1115—8.

[49] Beard DJ, et al. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. Lancet 2018;391:329—38.

[50] Abraha I, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. BMJ 2015;350:h2445.

[51] Hall M, et al. An evaluation of sex- and gender-based analyses in oncology clinical trials. J Natl Cancer Inst 2022. Aug 8;114(8):1186—91.

[52] White J, et al. The integration of sex and gender considerations into biomedical research: lessons from international funding agencies. J Clin Endocrinol Metab 2021;106:3034—48.

[53] Ranstam J, Cook JA. Kaplan-Meier curve. Br J Surg 2017;104:442.

[54] Castañon E, et al. Critical reappraisal of phase III trials with immune checkpoint inhibitors in non-proportional hazards settings. Eur J Cancer 2020;136:159—68.

[55] Dormuth I, et al. Which test for crossing survival curves? A user's guideline. BMC Med Res Methodol 2022;22:34.

[56] Dwan K, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PLoS One 2013;8:e66844.

[57] Andrade C. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices J Clin Psychiatry 2021;82.

[58] Rubin R. Mesh implants for women: scandal or standard of care? JAMA 2019;321:1338—40.

[59] Gluck O, et al. Implant files: news and paradox. J Gynecol Obstet Hum Reprod 2019;48:703—4.

[60] Buxton M. The economic evaluation of high technology medicine: the case of heart transplants. Bristol, 1986. In: Williams A, editor. Health and economics: proceedings of section F (economics) of the British association for the advancement of science. City: Palgrave Macmillan UK; 1987. p. 162–72.

[61] McCulloch P, et al. Progress in clinical research in surgery and IDEAL. Lancet 2018;392:88–94.

[62] Byrne BE, et al. A mixed methods case study investigating how randomised controlled trials (RCTs) are reported, understood and interpreted in practice. BMC Med Res Methodol 2020;20:112.

[63] Zhang Y, Guan X. Misleading reporting in statistically not significant oncology trials—joining efforts toward unbiased results interpretation. JAMA Netw Open 2021;4. e2138695-e.

[64] Strobel O, Büchler MW. The problem of the poor control arm in surgical randomized controlled trials. Br J Surg 2013;100:172–3.

[65] Cook AJ, et al. Statistical lessons learned for designing cluster randomized pragmatic clinical trials from the NIH Health Care Systems Collaboratory Biostatistics and Design Core. Clin Trials 2016;13:504–12.

[66] Haugen AS, et al. Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. Ann Surg 2015;261:821–8.

[67] Schumacher JR, et al. Increasing socioeconomically disadvantaged patients' engagement in breast cancer surgery decision-making through a shared decision-making intervention (A231701CD): protocol for a cluster randomised clinical trial. BMJ Open 2022;12:e063895.

[68] Martin J, et al. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. BMJ Open 2016;6:e010166.

[69] Simon GE, et al. Zelen design clinical trials: why, when, and how. Trials 2021;22:541.

[70] Feng K, et al. A randomized controlled trial of radiofrequency ablation and surgical resection in the treatment of small hepatocellular carcinoma. J Hepatol 2012;57:794–802.

[71] Wasmann KA, et al. Partially randomised patient preference trials as an alternative design to randomised controlled trials: systematic review and meta-analyses. BMJ Open 2019;9:e031151.

[72] Bach SP. Can we Save the rectum by watchful waiting or TransAnal surgery following (chemo)Radiotherapy versus Total mesorectal excision for early REctal Cancer (STAR-TREC)? Protocol for the international, multicentre, rolling phase II/III partially randomized patient preference trial evaluating long-course concurrent chemoradiotherapy versus short-course radiotherapy organ preservation approaches. Colorectal Dis 2022 May;24(5):639–51.

[73] Cook JA, et al. Surgeons' and methodologists' perceptions of utilising an expertise-based randomised controlled trial design: a qualitative study. Trials 2018;19:478.

[74] Nickel F, et al. Minimally Invasive versus open AbdominoThoracic Esophagectomy for esophageal carcinoma (MIVATE) - study protocol for a randomized controlled trial DRKS00016773. Trials 2021;22:41.

[75] Zolin SJ, et al. Registry-based randomized controlled trials: a new paradigm for surgical research. J Surg Res 2020;255:428–35.

[76] Foroughi S, et al. Re-inventing the randomized controlled trial in medical oncology: the registry-based trial. Asia Pac J Clin Oncol 2018;14:365–73.

[77] Ecker BL, et al. Registry-based randomized clinical trials in surgery: working with ACS-NSQIP and the AHPBA to conduct pragmatic trials. J Surg Oncol 2022;125:89–92.

[78] Foroughi S, et al. Registry-based randomized clinical trials as a method to improve cancer care in Australia. Asia Pac J Clin Oncol 2019;15:188–9.

[79] Karanatsios B, et al. Defining key design elements of registry-based randomised controlled trials: a scoping review. Trials 2020;21:552.

[80] Prang KH, et al. Clinical registries data quality attributes to support registry-based randomised controlled trials: a scoping review. Contemp Clin Trials 2022;119:106843.

[81] Fröbert O, et al. Thrombus Aspiration in ST-Elevation myocardial infarction in Scandinavia (TASTE trial). A multicenter, prospective, randomized, controlled clinical registry trial based on the Swedish angiography and angioplasty registry (SCAAR) platform. Study design and rationale. Am Heart J 2010;160: 1042–8.

[82] Renz BW, et al. Pylorus resection versus pylorus preservation in pancreatoduodenectomy (PyloResPres): study protocol and statistical analysis plan for a German multicentre, single-blind, surgical, registry-based randomised controlled trial. BMJ Open 2021;11:e056191.

[83] Crona J, et al. ENSAT registry-based randomized clinical trials for adrenocortical carcinoma. Eur J Endocrinol 2021;184. R51-r9.

[84] Schraa SJ, et al. Circulating tumor DNA guided adjuvant chemotherapy in stage II colon cancer (MEDOCC-CrEATE): study protocol for a trial within a cohort study. BMC Cancer 2020;20:790.

[85] Thorlund K, et al. Key design considerations for adaptive clinical trials: a primer for clinicians. BMJ 2018;360:k698.

[86] VanBuren JM, et al. The design of a Bayesian adaptive clinical trial of tranexamic acid in severely injured children. Trials 2021;22:769.

[87] Bhatt DL, Mehta C. Adaptive designs for clinical trials. N Engl J Med 2016;375: 65–74.

[88] Park JJH, et al. An overview of precision oncology basket and umbrella trials for clinicians. CA A Cancer J Clin 2020;70:125–37.

[89] Dreyer SB, et al. Precision oncology in surgery: patient selection for operable pancreatic cancer. Ann Surg 2020;272:366–76.

[90] Wendler T, et al. How molecular imaging will enable robotic precision surgery. Eur J Nucl Med Mol Imag 2021;48:4201–24.

[91] Kolla L, et al. The case for AI-driven cancer clinical trials - the efficacy arm in silico. Biochim Biophys Acta Rev Cancer 2021;1876:188572.