Expertise-based design in surgical trials: a narrative review

Ali Alsagheir, MBBS, MSc*
Alex Koziarz, MSc*
Emilie P. Belley-Côté, MD, PhD
Richard P. Whitlock, MD, PhD

*These authors contributed equally to this work.

Correspondence to:
R.P. Whitlock
David Braley Cardiac, Vascular and Stroke Research Institute
Room 1C1-5B
237 Barton St. E.
Hamilton ON L8L 2X2
richard.whitlock@phri.ca

Randomized controlled trials (RCTs) are the most robust study design for evaluating the safety and efficacy of a therapeutic intervention. However, their internal validity are at risk when evaluating surgical interventions. This review summarizes existing expertise-based trials in surgery and related methodological concepts to guide surgeons performing this work. We provide caseloads required to reach the learning curve for various surgical interventions and report criteria for expertise from published and unpublished expertise-based trials. In addition, we review design and implementation concepts of expertise-based trials, including recruitment of surgeons, crossover, ethics, generalizability, sample size and definitions for learning curve. Several RCTs have used an expertise-based design. We found that the majority of definitions used for expertise were vague, heterogeneous, and inconsistent across trials evaluating the same surgical intervention. Statistical methods exist to adjust for the learning curve; however, there is limited guidance. We developed the following criteria for surgical expertise for future trials: 1) decide on the proxy to be used for the learning curve, and 2) assess eligible surgeons by comparing their performance to the previously defined expertise criteria.

L'essai randomisé et contrôlé (ERC) représente le modèle d'étude le plus solide pour évaluer l'innocuité et l'efficacité d'une intervention, mais sa validité interne est compromise lorsqu'on évalue des interventions chirurgicales. Cette revue résume les essais existants sur l'expertise en chirurgie et les concepts méthodologiques connexes pour guider les chirurgiens dans leur travail. Nous donnons les volumes de cas requis pour atteindre la courbe d'apprentissage propre à diverses interventions chirurgicales et nous citons les critères d'expertise mentionnés dans les rapports d'essais sur l'expertise publiés et non publiés. Nous passons également en revue les protocoles et les concepts de mises en œuvre des essais sur l'expertise, y compris le recrutement des chirurgiens, la permutation des groupes, l'éthique, la généralisabilité, la taille des échantillons et les définitions des courbes d'apprentissage. Plusieurs ERC ont utilisé un modèle basé sur l'expertise. Nous avons découvert que la majorité des définitions du terme expertise utilisées étaient vagues, hétérogènes et inconstantes d'un essai à l'autre pour une même intervention chirurgicale. Il existe des méthodes statistiques pour ajuster la courbe d'apprentissage; par contre, on déplore un manque de directives. Nous avons établi les critères suivants pour juger de l'expertise chirurgicale en vue de futurs essais : 1) choisir la mesure indirecte à utiliser pour la courbe d'apprentissage, et 2) évaluer les chirurgiens admissibles en comparant leur rendement aux critères d'expertise définis précédemment.

Randomized controlled trials (RCTs) are the most robust study design for assessing an intervention's efficacy and safety. However, when assessing skill-dependent interventions such as surgical interventions, their internal validity can be compromised by performance bias. The characteristics of the participating surgeons and centres can affect study outcomes and therefore validity. Surgeons generally favour one technique over others and become less familiar with alternative operative approaches. Requesting that they perform procedures with which they are less comfortable in the context of a trial may lead to biased and misleading results.[1–4] To minimize this problem, expertise-based trials are an alternative to conventional RCTs.[1–5]

Surgeons are an increasingly common and critical resource when evaluating procedural interventions with RCTs. Therefore, we summarized existing trials in the surgical literature from an expertise-based lens, providing a summative resource to guide future surgeons leading this work. We examined methodological concepts inherent to the design and conduct of expertise-based trials. Finally, we developed criteria for surgical expertise for the design of future trials.

Definition of expertise-based trials

In a conventional RCT, participants are randomly allocated to either the intervention or control arm, and health care providers provide both interventions. However, in an expertise-based trial, the health care provider will vary based on allocation owing to their expertise for the control or experimental arm to ensure that an expert performs the allocated intervention.[3,5] This minimizes the impact of clinician experience on patient outcomes and strengthens internal validity. Expertise-based design is therefore increasingly adopted to compare skill-based interventions, even within the same field.[6,7]

Importance of expertise-based trials in surgery

In surgical trials, intervention-related factors that may affect outcome include the procedure, the surgeon, the surgical team, and pre- and postoperative care.[8,9] In RCTs evaluating surgical interventions, these components should be balanced to maintain the advantages of randomization and to infer causality. Conventional surgical RCTs are bound to be criticized when surgeons provide both interventions without considering their expertise.[10,11]

METHODS

We performed a literature search of MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials from inception to May 2018 for expertise-based trials in surgery. The following keywords were used: "randomi?ed," "RCT," "expertise," "learning curve," "random," "surgical," "operative," and "procedure." Studies that included procedures delivered by clinicians or interventionists (e.g., interventional cardiologists, gastroenterologists, interventional radiologists) were included. Expertise-based trials and their definitions of expertise were reported qualitatively. Caseloads used for expertise for specific procedures, such as off-pump coronary artery bypass graft (CABG), were also reported. Methodology papers were critically summarized. A team including cardiac surgeons, a cardiologist, and methodologist summarized the current methodological literature on expertise-based trials. In addition, we developed criteria for surgical expertise as a suggestion for the design of future expertise-based trials.

RESULTS

Review of methodological literature — potential advantages of expertise-based trials

Differential expertise bias

Surgical experience and comfort in performing an intervention may influence the estimate of effect. This may be negligible if the variation in technique is small or if the "new" intervention requires limited procedural skills. However, expertise becomes key when the interventions vary significantly, making practice and familiarity crucial to procedural outcomes. In general, surgeons preferentially employ a single approach to treat a specific problem.[4,12] The consequent lack of familiarity with one of the interventions and expertise in the other could lead to differential expertise bias.[1–4] If surgeons participating in a trial are more comfortable and experienced with intervention A than intervention B, we expect the estimated effect size to be falsely skewed in favour of intervention A. For all surgical procedures, training is associated with a learning curve as proficiency develops. Depending on the complexity of the procedure and the individual surgeon's skills, the number of cases required to achieve competence may vary substantially.[3,13] For aortic or mitral valve surgery, surgeon experience correlates with reduced cardiopulmonary bypass time and improved long-term survival, suggesting that expertise contributes to procedural outcomes.[6] However, for on-pump CABG, surgeon experience does not correlate with similar benefits;[6,7,14] this could be attributable to the higher frequency of CABG surgery, leading the learning curve for CABG to occur during training and making all surgeons experts at this procedure. Accordingly, effect sizes produced by conventional surgical intervention trials may be misleading if investigators do not account for surgeon expertise.

Improved recruitment of surgeons

Conventional RCT designs require surgeons to treat patients with procedures that may deviate from their personal preference, potentially reducing surgeon buy-in and slowing trial recruitment. Studies assessing the impact of expertise-based designs on surgeons' willingness to participate have shown mixed results. In a survey of orthopedic surgeons asked to participate in a trial comparing high tibial osteotomy and unicompartmental knee arthroplasty,[8,15] 53% were willing to participate in an expertise-based design, but only 18% in a conventional RCT. Another survey assessing the willingness of Canadian vascular surgeons to participate in a trial comparing endovascular aneurysm repair (EVAR) and open aortic repair did not find a significant difference in design preference.[10,16] These divergent results suggest that the response to

expertise-based design is context specific. Another plausible explanation includes the types of clinical equipoise involved. If the clinician is truly uncertain and has no preference for one intervention over another, they may be more likely to participate in a conventional RCT. Alternatively, if there is disagreement regarding the efficacy or safety of 2 interventions between 2 groups of clinicians, there may be less participation in a conventional RCT. Surveying physician perspectives at the time of trial development may inform design decisions.

Reduced crossover

High crossover resulting from surgeon bias or procedural preference can compromise the effect estimate. In expertise-based designs, surgeons are more likely to adhere to the assigned intervention as they may believe in its superiority, may be more comfortable performing it, or may be more experienced in general.[2,3,17] Ideally, in an expertise-based design, crossovers should occur only for absolute anatomic requirements or intraoperative complications, and never because of physician discomfort with the allocated procedure. To examine the potential advantage of expertise-based design with respect to crossover rate, we examined 2 RCTs evaluating CABG technique (on- v. off-pump): 1 with expertise-based design (CORONARY trial)[7,14,18] and 1 conventional RCT (ROOBY trial).[5,15] There was no predefined specification of expertise in the ROOBY trial, with both residents and attending surgeons performing the operations.[15] In the CORONARY trial, primary surgeons were required to have more than 2 years of postresidency experience in the procedure and have performed more than 100 cases.[15] In the ROOBY trial, there was a 12.4% crossover rate from off- to on-pump, compared with 7.9% in the CORONARY trial ($p < 0.001$). In the ROOBY trial, relative risk of death at 5 years was 1.28 (95% confidence interval [CI] 1.03–1.58, $p = 0.03$),[15] whereas in the CORONARY trial the hazard ratio of death was 1.08 (95% CI 0.93–1.26, $p = 0.30$).[7] In a systematic review of expertise-based trials, 92% of participants received the allocated treatment (interquartile range [IQR] 82%–99%).[5,16]

Ethics

Requiring that surgeons perform procedures that they are unfamiliar or uncomfortable with introduces unnecessary risks, which are difficult to justify ethically. In an expertise-based trial, experts deliver surgical interventions, limiting risk and providing a "real-life" setting for trial participants. Expertise-based designs are ethically reasonable: patients undergo routine procedures performed at a high competence level by an experienced surgeon. This may also facilitate research ethics board approval as the studied interventions more closely fall within standard of care.

Potential disadvantages of expertise-based trials

Reduced generalizability

Although expertise-based designs are often considered pragmatic, their results can only be generalized to expert surgeons or centres. The generalizability of an expertise-based trial depends on the participating surgeons or centres and the scope of the trial. Surgeons have to meet the expertise requirements to be expected to achieve similar results. Criteria for expertise can be adjusted to increase the applicability of the results to the surgical community.

Increased sample size

Calculating the required sample size and ensuring adequate patient recruitment is crucial to generate robust conclusions.[1,19] In surgical trials, participants treated by the same surgeon are more likely to have similar outcomes than those under the care of another surgeon.[20] This cluster effect has direct implications on sample size calculations. In typical expertise-based trials, surgeons perform only 1 intervention, which leads to a higher cluster effect and therefore standard error than in conventional RCTs where surgeons can perform both interventions under investigation.[2,21] This increase in standard error with expertise-based designs leads to a larger required sample size than in conventional trials. Relative efficiency is a ratio of standard errors for the estimated effects of the expertise-based compared with the conventional design (Figure 1). Assuming the number of surgeons and patients are the same in both designs, the efficiency will depend on the number of patients treated by each surgeon in each treatment arm, variance of surgeon effect, and variance of treatment effect. Based on the formula, the efficiency of expertise-based designs can be increased by using more surgeons, with fewer patients per surgeon.[2]

In expertise-based designs, patient outcomes and observed treatment effects may affect the sample size owing to the expertise of participating surgeons. First, one can anticipate lower overall event rates of safety outcomes owing to the expertise of participating clinicians, which increases sample size.[2] Second, a reduced crossover rate between interventions retains statistical power in an intention-to-treat analysis that would otherwise be reduced with increased crossover rate, which is more common in conventional trials.[22]

$$RE = \frac{var(\theta)EB}{var(\theta)C} = \frac{m'k'}{mk}\left[1 + \frac{m\sigma_s^2}{\sigma^2}\right] = 1 + m\sigma_s^2/\sigma^2$$

Fig. 1. Relative efficiency of standard errors for estimated effects of expertise-based versus conventional design. Variables: m denotes number of patients treated by each surgeon in each treatment arm, k denotes number of surgeons, EB denotes expertise-based, C denotes conventional, and σ denotes variance.

Logistical issues

The coordination of expertise-based trials is challenging. These trials require experts in both studied interventions to be available, which may be demanding for emergency procedures. Timing of randomization also poses a challenge. If a surgeon who is an expert in only 1 of the procedures recruits participants, patients may be reluctant to change surgeon. Recruitment through a third party, such as a resident, emergency department physician, primary care physician, or referring physician, before the surgical consultation may avoid this problem. Another logistical issue arises at the patient level if a centre does not offer both interventions. Centres with expertise in only 1 intervention must be willing to transfer the patients to other centres if they are allocated to the other intervention. Patients must also agree to possible transfer or travel based on trial allocation, which may affect outcomes or be burdensome to patients.[3,11]

Definitions of expertise in the literature

The lack of standardized definition of expertise becomes a fundamental issue in designing expertise-based trials. Expertise criteria are procedure-specific and aim to address issues around the learning curve. Number of procedures performed has initially been a common proxy to define expertise, with several studies in cardiac surgery reporting cutoffs for various procedures (Table 1). However, expertise is also associated with professional grade, years of experience, and annual caseload.

Learning curve

A learning curve is "the time taken and/or the number of procedures an average surgeon needs to be able to perform a procedure independently with a reasonable outcome."[18] All procedures have a learning curve that correlates with procedure complexity. Cook and colleagues[5] proposed 3 phases to characterize the learning curve: the starting point, where the surgeon begins performing a specific intervention; the rate of learning, which reflects the speed of skill acquisition for a surgeon to reach a certain level of performance; and the expert level or asymptote, where the surgeon's performance stabilizes. During the early phases of the learning curve, errors and adverse outcomes are more likely to occur.

The steepness of the learning curve depends on the type of procedure; outcome measures; level of prior experience; and the surgeon's annual caseload, institution, and inherent skill. It has been suggested that a minimum of 10 procedures is required to reach the learning curve asymptote for a procedural intervention.[5] However, in our view, this is too simplistic given the spectrum of intervention complexity. For example, the learning curve of coronary artery

bypass surgery was reported to be 15–100 cases, depending on procedure type and experience. Table 2 provides a summary of surgical expertise-based trials in different surgical specialties with their expertise definitions. Commonly reported proxies for expertise included number of years in practice, total number of cases performed and number of cases performed per year. Number of cases performed per year without accounting for number of years in practice may be limited, as surgeons with more years in practice are more likely to be at an asymptote in their learning curve. One trial used a proxy of high procedural success and low complication rate;[37] however, this may limit trial participation to surgeons who restrict their practice caseload to lower-risk patients, which may inflate their expertise level. Accordingly, number of years in practice as a primary surgeon and total number of cases may be appropriate proxies. Table 3 provides several examples of commonly used proxies for expertise and reports their respective rationale.

Surgeons learn at different speeds, and the learning process depends on external factors, such as previous expertise in similar procedures. Therefore, the performance of a specific number of operations does not guarantee that the technical asymptote is met. More rigorous standards to demonstrate surgical competence have been proposed, such as direct training and supervision, specifying an outcome performance level, recommendation by experts, and assessing performance/outcome. However, none of these approaches is well studied, and the understanding of the learning process remains incomplete.

Statistical methods and issues to account for learning curve

Several statistical approaches have been described to adjust for the operator learning curve when evaluating surgical trials. These approaches range from descriptive to more complex, including split group, univariate analysis, multivariate analysis, cumulative sum, and multilevel analysis.[1] Guidance on the statistical model to use is lacking.[21,38] Multilevel modelling is a promising method because it adjusts for different operators and institutions and considers case mix; however, it generally requires a larger sample size.[3,5,38]

The main issue related to identifying the expertlevel phase in a learning curve is the scarcity and/or

Table 1. Caseloads used for various cardiac surgeries in expertise-based trials	
Procedures	No. of cases to reach learning curve
Off-pump CABG[7]	85
Minimally invasive CABG[23]	40
Robotic-assisted left internal mammary harvest[24]	15
Off-pump endoscopic CABG[25]	> 100
CABG = coronary artery bypass graft.	

Trial	Specialty	Intervention	No. of participants and surgeons	Criteria for expertise
Table 2. Expertise-based randomized controlled trials identified after systematic literature search of MEDLINE and EMBASE				
Finkemeier et al.[29]	Orthopedics	Nail insertion with reaming v. without reaming	94 participants 6 surgeons	NR
Phillips et al.[30]	Orthopedics	Open reduction and internal fixation v. closed cast treatment	138 participants 2 surgeons	NR
Machler et al.[31]	Cardiac	Mini aortic valve surgery (2 surgeons) v. conventional aortic valve (2 surgeons)	120 participants 4 surgeons	NR
Wihlborg et al.[26]	Orthopedics	Rydell 4-flanged nail v. Gouffon pins	200 participants 7 surgeons	• No. of years of training and practice
Wyrsch et al.[32]	Orthopedics	ORIF of the tibia and fibula v. external fixation with or without limited internal fixation	39 participants 6 surgeons	• Fellowship-trained trauma surgeons • Experts
CABRI[27]	Cardiac surgery/ cardiology	Coronary angioplasty v. CABG	1054 participants	• High-volume centres • Surgeon and physician performed > 500 procedures
RITA trial[33]	Cardiac surgery/ cardiology	Coronary angioplasty v. CABG	1011 participants	• High success and low complication rate
BARI[28]	Cardiac surgery/ cardiology	Coronary angioplasty v. CABG	1829 participants	• Yr of practice (3 yr) • Outcome assessment "Majority of practice devoted to coronary artery surgery; most recent 100 consecutive primary, elective, isolated CABG operations with a mortality rate of no more than 2% (death within 30 days of procedure); and an MI rate of no more than 4%" • No. of cases "as principal surgeon of 100 or more CABGs with internal mammary artery grafts" • Number of cases "independent operator in more than 300 elective PTCA procedures, of which 100 were multivessel disease cases"
CORONARY[23]	Cardiac surgery	Off-pump v. On-pump CABG	4752 participants	• Yr of practice (2 yr) • No. of cases "more than 100 procedures involving the specific technique"
TOPKAT[34] (ongoing) NCT01352247	Orthopedics	Total v. partial arthroplasty	500 participants	• "Simple audit of participating surgeons' routine practice will be undertaken" • Appropriate training • Yr of experience "at least 1 yr" • No. of cases "have performed the operation at least 10 times in the past year"
HEALTH[35] (ongoing) NCT00556842	Orthopedics	THA v. HA	1501 participants	• No. of cases "Perform at least 50 procedures (either THA or HA) in their career" • Annual load "at least 5 procedures per year"
NExT ERA[36] (ongoing) NCT00358085	Vascular	Open v. endovascular repair for AAA	30 participants	• Appropriate training "completion of a vascular residency at a credentialed academic centre, a period of study in a formal training program dedicated to acquiring endovascular expertise" or for open AAA "require completion of an accredited vascular surgery residency program • No. of cases "experience with at least 60 previous EVAR procedures" or "at least 100 consecutive elective AAA repairs"

AAA = abdominal aortic aneurysm; CABG = coronary artery bypass grafting; EVAR = endovascular aneurysm repair; HA = hemiarthroplasty; MI = myocardial infarction; NR = not reported; ORIF = open reduction internal fixation; PTCA = percutaneous transluminal coronary angioplasty; THA = total hip arthroplasty.

poor quality of available quantitative data in the literature. In a systematic review of statistical methods used to assess surgeons' learning curve, most of the data came from descriptive methods that are considered exploratory analyses. These analyses identify the existence of the learning process but cannot estimate the parameters of the learning curve.[1,3,5] Furthermore, most of these studies were based on case series, and 64% addressed the learning curve for a single operator, which limits the generalizability of the estimated learning points to other surgeons.[1,39,40] In addition, some surgeons reach a plateau phase in their learning curve which is lower than an agreed upon expertise level.[7,14] For this reason, a performance criterion may be needed in the expertise definition to ensure capturing only expert surgeons.

Table 3. General proxies for expertise in RCTs	
Expertise	Rationale
Proxies for learning curve	
Total no. of cases	To reach learning curve plateau
Annual case volume	Volume necessary to maintain proficiency
Yrs of independent practice	Prior experience with other type of surgeries could shorten the learning curve
Previous training	Affects the slope of learning curve
Competence	
Supervision before participation (e.g., Fellowship, extra training)	Indirect indication of competence
Performance/outcome assessment (e.g., audit)	Indication that competence and the learning plateau match the study's predefined expertise levels
Recommendation by expert	Indirect indication of competence
RCT = randomized controlled trial.	

Review of included expertise-based RCTs

There is considerable heterogeneity in the extent of reporting and requirements for surgeon participation in expertise-based trials. Wihlborg[26] reported an expertise-based RCT including 200 patients, 7 surgeons, and 1 centre to evaluate the use of Rydell 4-flanged nail compared with Gouffon pins for cervical hip fractures. Instead of predefining criteria for expertise, orthopedic surgery groups were selected with similar number of years of training and practice with each surgical technique. This methodology may be practical for trials including few centres and interventions with similar learning curves; however, its use many lead to differential expertise bias when comparing interventions with different learning curves. Importantly, this trial was conducted between 1984 and 1987, several decades before the emergence of clinicians and surgeons calling for increased use of expertise-based design.[3]

The CABRI (coronary angioplasty versus bypass revascularisation investigation) study[27] was an expertise-based trial including 1054 patients and 26 centres evaluating CABG and percutaneous transluminal coronary angioplasty (PTCA) in patients with symptomatic multivessel coronary disease. Participating centres were required to be high volume, defined as "each principal surgeon and physician having performed at least 500 procedures."[27] The crossover rate of CABG was 3.9%, whereas for PTCA it was 2.8%. No further details regarding expertise were provided, nor did any analyses examine the effect of potential differential expertise on patient outcomes. Accordingly, replicating this trial may become challenging, particularly in different health care settings.

The BARI (bypass angioplasty revascularization investigation) study[28] was an expertise-based trial evaluating CABG compared with PTCA in advanced coronary artery disease with several requirements from surgeons and clinicians to participate. Clinicians performing PTCA were required to have a lesion success rate higher than 85% for subtotal lesions among the last 100 cases, an overall incidence per patient of PTCA-related acute myocardial infarction or emergency CABG of 5% or less, and a mortality rate for elective cases less than 2%.[28] Surgeons performing CABG were required to have a mortality rate less than 2% and a myocardial infarction rate less than 4% for the most recent 100 consecutive cases.[28] Whereas stringent eligibility criteria for clinicians and surgeons increase internal validity of the trial, the external validity of the trial is reduced for surgeons who may have differing levels of expertise or manage more complex cases with higher risks for morbidity and mortality.

The CORONARY trial[23] was an expertise-based based trial of 4752 patients, 79 centres, and 19 countries evaluating off-pump compared with on-pump CABG. Expertise was defined as nontrainees with more than 2 years of experience after residency training and completion of more than 100 cases of off-pump or on-pump CABG.[23] The crossover rate was 7.9% for patients randomized to off-pump and 6.4% for patients randomized to on-pump surgery ($p = 0.06$).[23] The most common reason for crossover in the off-pump group was hypotension (31.2%), followed by small targets (25.5%), intramuscular vessels (22.3%), and ischemia (18.5%).[23] The most common reason for crossover in the on-pump group was calcified ascending aorta (64.7%), followed by patient comorbidities (7.7%).[23] Similar to the BARI trial,[28] the reporting of criteria used for expertise were detailed, allowing for trial replication and future evaluation of the impact of variable levels of operator expertise on patient outcomes. Trials with notably high crossover rates with recurrent reasons for crossover (e.g., common anatomic variations hindering conduction of a procedure) highlight the importance of specifying eligibility criteria whereby included patients would be candidates for both procedures.

DISCUSSION

A suggested approach to expertise definition

To develop expertise criteria for a specific intervention, 2 steps should be completed (Table 4):

1. Decide on the proxy to be used for the learning curve. This may include number of years in practice, number of previous cases performed, annual caseload, and training. Within that proxy, establish a consensus of the requirement to reach the plateau of the learning curve.
2. Assess eligible surgeons by comparing their performance with the previously defined expertise criteria. If the expertise of a surgeon is near the predefined proxy cutoff, they may increase their expertise to meet the requirements and participate in the trial after another expertise evaluation.

Table 4. Proposed proxy for evaluating expertise in surgical trials	
Expertise definition	Proposed method for evaluating expertise
The level of proficiency of a surgeon whereby one is able to perform a procedure independently with reasonable outcomes consistently	1) Decide on the proxy to be used for the learning curve 2) Assess eligible surgeons by comparing their performance with the previously defined expertise criteria

To address the first point, investigators can review previously reported learning curve estimates for the interventions of interest (or similar interventions) and consider the methods used for evaluating the learning curve. In scenarios where previous reports are either not robust or lacking, conducting a survey of experts in the field to explore the expected amount of experience required may provide a reliable alternative.

When assessing competence, it is important to ensure that surgeons who are clearly observed to have reached the learning plateau fulfill the study's expertise criteria. If used, performance measurements should be based on outcomes adjusted for preoperative risk;[9,14] expert surgeons may operate on higher-risk patients, and raw performance measures may be misleading. Other methods that can be used include providing evidence of sufficient training, supervision before participating and recommendation by an expert.

Survey evaluation of criteria for both feasibility and appropriateness by key opinion leaders could add face validity, and criteria may be optimized by direct community feedback. The trial's perspective can also influence criteria for expertise: more lenient criteria would be in line with a pragmatic trial. Meanwhile, an explanatory trial would call for stricter criteria, including only highly experienced operators allowing for the evaluation of the intervention's efficacy and safety under ideal conditions.

Centres involved in expertise-based trials ideally must have expertise in both interventions of interest. When planning an expertise-based trial, a feasibility survey may help to identify eligible centres based on the number of health providers who are experts in each technique, the willingness of health providers to participate in an expertise-based trial, and the annual number of eligible participants for an RCT at each centre.

Trials that use expertise-based design should unambiguously report the criteria used to define expertise and provide appropriate justification. Failure to do so can lead to criticism and decrease the applicability of the trial's findings. In general, surgical trials, whether expertise-based or conventional, should aim for transparency when reporting on expertise. Future CONSORT (CONsolidated Standards of Reporting Trials) statements should address this issue.

Innovation, evaluation, and regulation of surgical therapies using expertise-based design

Pharmaceuticals in North America undergo a highly structured review by the US Food and Drug Administration or Health Canada that is split into several phases ranging from phase 0 (assess drug pharmacokinetics) to phase 4 (postmarketing surveillance of drug assessing rare adverse events). Unlike this regimented assessment of drugs in North America, the majority of surgical interventions remain unregulated.[24] The development of surgical interventions has been based on anatomic and pathophysiological principles, often facing several stages of improvement as they become adopted by other surgeons.[24] In fact, the innovation and adoption of surgical therapies has been described by the Balliol Collaboration as distinct stages:[25] stage 0, prehuman work to develop technique; stage 1, development of technical skills and evaluation of safety (may be described in case reports); stage 2a, adoption of technique by surgical leaders); stage 2b, technical details of procedure are approaching asymptote (may be described in prospective cohort studies or RCTs); stage 3, procedure is standard of practice for most surgeons; and stage 4, long-term surveillance studies implemented to monitor for rare outcomes.

Given the continual improvement in novel surgical procedures, it is often challenging to decide when to evaluate an innovative surgical procedure with an RCT. Implementing an RCT during an early stage of a novel surgical technique may be argued to reflect the surgeons not reaching a sufficient level of the procedure's learning curve, or the procedure still being in its early stages of innovation.[41] Consistent monitoring of the safety and efficacy during the early stages of innovative procedures must remain a priority, which may be collected using registries that include the procedure's successes and failures.[42] The decision to conduct an expertise-based trial may be most appropriate during early adoption of the surgical technique whereby several surgeons have learned the technique,[25,41] allowing for the development of a proxy for the learning curve.

CONCLUSION

In expertise-based trials, participants are randomized to interventions performed by physicians with expertise in the assigned intervention. This design addresses criticisms associated with conventional trials for surgical interventions, including differential expertise bias and learning curves. Other theoretical advantages are associated with expertise-based design such as limiting crossovers, minimizing bias related to blinding of surgeons, and increasing recruitment of patients and surgeons.

However, the main challenge with expertise-based design is how to define expertise. Because expertise cannot be measured directly, indirect proxies are typically used, and current approaches often lack justification. We proposed a general approach to define expertise assessing descriptions of procedure-specific learning curves and ensuring that the predefined expertise criteria are met. When identifying surgeons eligible for participation, researchers should aim to recruit surgeons in the final stage of the learning curve and to ensure they meet predefined expertise criteria. Moreover, completed studies should report on recruited surgeons' experience and levels of expertise to facilitate interpretation and generalization of results. Another potential disadvantage of expertise-based trials is the impact of clustering around surgeons on sample size calculations and statistical analysis.

Expertise-based trials are a new design, well suited to study surgical interventions. As the use of expertise-based design increases, our understanding of its challenges and potential solutions will evolve.

Affiliations: From the Division of Cardiac Surgery, Department of Surgery, McMaster University, Hamilton, Ont. (Alsagheir, Whitlock); the Faculty of Medicine, University of Toronto, Toronto, Ont. (Koziarz); the Population Health Research Institute, McMaster University and Hamilton Health Sciences, Hamilton, Ont. (Belley-Côté, Whitlock); and the Department of Medicine, McMaster University, Hamilton, Ont. (Belley-Côté).

Competing interests: None declared.

Contributors: All authors designed the study. A. Alsagheir and A. Koziarz acquired the data, which all authors analysed. A. Alsagheir and A. Koziarz wrote the article, which all authors reviewed. All authors approved the final version to be published.

References

1. Ramsay CR, Grant AM, Wallace SA, et al. Assessment of the learning curve in health technologies. A systematic review. *Int J Technol Assess Health Care* 2000;16:1095-108.
2. Walter SD, Ismaila AS, Devereaux PJ, et al. Statistical issues in the design and analysis of expertise-based randomized clinical trials. *Stat Med* 2008;27:6583-96.
3. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ* 2005;330:88.
4. van der Linden W. Pitfalls in randomized surgical trials. *Surgery* 1980;87:258-62.
5. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clin Trials* 2004;1:421-7.
6. Burt BM, ElBardissi AW, Huckman RS, et al. Influence of experience and the surgical learning curve on long-term patient outcomes in cardiac surgery. *J Thorac Cardiovasc Surg* 2015;150:1061-7.
7. Lamy A, Devereaux PJ, Yusuf S. Five-year outcomes after off-pump or on-pump coronary-artery bypass grafting. *N Engl J Med* 2017;376:894-5.
8. Bednarska E, Bryant D, Devereaux PJ; Expertise-Based Working Group. Orthopaedic surgeons prefer to participate in expertise-based randomized trials. *Clin Orthop Relat Res* 2008;466:1734-44.
9. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials* 2009;10:9.
10. Mastracci TM, Clase CM, Devereaux PJ, et al. Open versus endovascular repair of abdominal aortic aneurysm: a survey of Canadian vascular surgeons. *Can J Surg* 2008;51:142-8.
11. Farrokhyar F, Karanicolas PJ, Thoma A, et al. Randomized controlled trials of surgical interventions. *Ann Surg* 2010;251:409-16.
12. Edward SJL, Stevens AJ, Braunholtz DA, et al. The ethics of placebo-controlled trials: a comparison of inert and active placebo controls. *World J Surg* 2005;29:610-4.
13. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
14. Papachristofi O, Jenkins D, Sharples LD. Assessment of learning curves in complex surgical interventions: a consecutive case-series study. *Trials* 2016;17:266.
15. Shroyer AL, Hattler B, Wagner TH, et al. Five-year outcomes after on-pump and off-pump coronary-artery bypass. *N Engl J Med* 2017;377:623-32.
16. Cook JA, Elders A, Boachie C, et al. A systematic review of the use of an expertise-based randomised controlled trial design. *Trials* 2015;16:241.
17. Lobato AC, Rodriguez-Lopez J, Diethrich EB. Learning curve for endovascular abdominal aortic aneurysm repair: evaluation of a 277-patient single-center experience. *J Endovasc Ther* 2002;9:262-8.
18. Subramonian K, Muir G. The 'learning curve' in surgery: what is it, how do we measure it and can we influence it? *BJU Int* 2004;93:1173-4.
19. Karlsson J, Engebretsen L, Dainty K; ISAKOS Scientific Committee. Considerations on sample size and power calculations in randomized clinical trials. *Arthroscopy* 2003;19:997-9.
20. Cook JA, Bruckner T, MacLennan GS, et al. Clustering in surgical trials-database of intracluster correlations. *Trials* 2012;13:2.
21. Khan N, Abboudi H, Khan MS, et al. Measuring the surgical 'learning curve': methods, variables and competency. *BJU Int* 2014;113:504-8.
22. Weinstein GS, Levin B. Effect of crossover on the statistical power of randomized studies. *Ann Thorac Surg* 1989;48:490-5.
23. Lamy A, Devereaux PJ, Prabhakaran D, et al. Five-year outcomes after off-pump or on-pump coronary-artery bypass grafting. *N Engl J Med* 2016;375:2359-68.
24. Roberts DJ, Zygun DA, Ball CG, et al. Challenges and potential solutions to the evaluation, monitoring, and regulation of surgical innovations. *BMC Surg* 2019;19:119.
25. Barkun JS, Aronson JK, Feldman LS, et al. Balliol Collaboration. Evaluation and stages of surgical innovations. *Lancet* 2009;374:1089-96.
26. Wihlborg O. Fixation of femoral neck fractures. A four-flanged nail versus threaded pins in 200 cases. *Acta Orthop Scand* 1990;61:415-8.
27. Participants CT. First-year results of CABRI (coronary angioplasty versus bypass revascularisation investigation). *Lancet* 1995;346:1179-84.
28. Szczech LA, Best PJ, Crowley E; Bypass Angioplasty Revascularization Investigation (BARI) Investigators. Outcomes of patients with chronic renal insufficiency in the bypass angioplasty revascularization investigation. *Circulation* 2002;105:2253-8.
29. Finkemeier CG, Schmidt AH, Kyle RF, et al. A prospective, randomized study of intramedullary nails inserted with and without reaming for the treatment of open and closed fractures of the tibial shaft. *J Orthop Trauma* 2000;14:187-93.
30. Phillips WA, Schwartz HS, Keller CS, et al. A prospective, randomized study of the management of severe ankle fractures. *J Bone Joint Surg Am* 1985;67:67-78.

31. Mächler HE, Bergmann P, Anelli-Monti M, et al. Minimally invasive versus conventional aortic valve operations: a prospective study in 120 patients. *Ann Thorac Surg* 1999;67:1001-5.

32. Wyrsch B, McFerran MA, McAndrew M, et al. Operative treatment of fractures of the tibial plafond. A randomized, prospective study. *J Bone Joint Surg Am* 1996;78:1646-57.

33. Participants RT. Coronary angioplasty versus coronary artery bypass surgery: the randomized intervention treatment of angina (RITA) trial. *Lancet* 1993;341:573-80.

34. Beard D, Price A, Cook J, et al. Total or partial knee arthroplasty trial — TOPKAT: study protocol for a randomised controlled trial. *Trials* 2013;14:292.

35. Bhandari M, Devereaux PJ, Einhorn TA, et al. Hip fracture evaluation with alternatives of total hip arthroplasty versus hemiarthroplasty (HEALTH): protocol for a multicentre randomised trial. *BMJ Open* 2015;5:e006263.

36. NExT ERA: national expertise based trial of elective repair of abdominal aortic aneurysms: a pilot study. ClinicalTrials.gov: NCT00358085; 2006. Available: https://clinicaltrials.gov/ct2/show/NCT00358085 (accessed 2020 Nov. 17).

37. Taggart DP. The RITA trial. *Lancet* 1993;341:1020.

38. Ramsay CR, Wallace SA, Garthwaite PH, et al. Assessing the learning curve effect in health technologies. Lessons from the nonclinical literature. *Int J Technol Assess Health Care* 2002;18:1-10.

39. Novick RJ, Fox SA, Stitt LW, et al. Assessing the learning curve in off-pump coronary artery surgery via CUSUM failure analysis. *Ann Thorac Surg* 2002;73:S358-62.

40. King SB, Lembo NJ, Weintraub WS, et al. A randomized trial comparing coronary angioplasty with coronary bypass surgery. Emory angioplasty versus surgery trial (EAST). *N Engl J Med* 1994;331:1044-50.

41. Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet* 2009;374:1097-104.

42. McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009;374:1105-12.