ORIGINAL ARTICLE

Funders improved the management of learning and clustering effects through design and analysis of randomized trials involving surgery

Elizabeth J. Conroy [a],[*], Anna Rosala-Hallas[a], Jane M. Blazeby[b], Girvan Burnside[c], Jonathan A. Cook[c], Carrol Gamble[a],[d]

[a]*Department of Biostatistics, University of Liverpool, A member of Liverpool Health Partners, Liverpool, UK*
[b]*Centre for Surgical Research, Population Health Sciences, University of Bristol, Bristol, UK*
[c]*Centre for Statistics in Medicine, University of Oxford, Oxford, UK*
[d]*North West Hub for Trials Methodology Research, University of Liverpool, Liverpool, UK*

Abstract

Objective: The objective of this study was to provide insight into current practice in planning for, and acknowledging, the presence of learning and clustering effects, by treating center and surgeon, when developing randomized surgical trials.

Study Design and Setting: Complexities associated with delivering surgical interventions, such as clustering effects, by center or surgeon, and surgical learning should be considered at trial design. Main trial publications, within the wider literature, under-report these considerations. Funded applications, within a 4-year period, from a leading UK funding body were searched. Data were extracted on considerations for learning and clustering effects and the driver, funder, or applicant, behind these.

Results: Fifty trials were eligible. Managing learning through establishing predefined center and surgeon credentials was common. One planned exploratory analysis of learning within center, and two within surgeon. Clustering, by site and surgeon, was often managed through stratifying randomization, with 81% and 60%, respectively, also planning to subsequently adjust analysis. One-third of responses to referees contained funder led changes accounting for learning and/or clustering.

Conclusion: This review indicates that researchers do consider impact of learning and clustering, by center and surgeon, during trial development. Furthermore, the funder is identified as a potential driver of considerations. © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Randomized controlled trials; Surgery; Clustering; Learning curve; Statistics; Surgical intervention

1. Introduction

Randomized controlled trials (RCTs) are recognized as providing the highest level of evidence, second only to systematic reviews of such trials [1]. The need for surgical randomized trials is well recognized [2,3], and this has led to a push for growth in recent years [3,4]. Leading research organizations are supporting this growth through establishing a number of initiatives and research objectives, ultimately aiming to improve the global surgical evidence base [5—10]. One such initiative, set up by the UK's leading publicly funded health research body, the National Institute

* Corresponding author. Medicines for Children Clinical Trials Unit, Clinical Trials Research Centre, University of Liverpool, Institute of Child Health, Alder Hey Children's NHS Foundation Trust, Liverpool, L12 2AP, UK. Tel.: +44 151 795 8791; fax: +44 151 795 8770.

E-mail address: ejconroy@liverpool.ac.uk (E.J. Conroy).

What is new?

Key findings
- Results indicate that while considerations of surgical learning and clustering are underreported in main trial publications, funders and researchers alike appear to be aware of the need to manage these aspects at trial design stage.

What this adds to what is known?
- A novel assessment of the decision making behind intended design and analysis with respect to the management of surgical learning and clustering is presented.

- This review investigates successful funding applications comprising a wide variety of trials, both by surgical discipline and by geographic location, by a leading UK funder.

- This review is timely as it comprises successful funding applications following a call by this funder recognizing a need for an increase in evidence-based surgical research.

What is the implication and what should change now?
- Insight into the promising role of the funder as a driver to improving the, long criticized, surgical evidence base is provided.

for Health Research (NIHR), aimed to increase the volume of high-quality research, across surgical disciplines, on the effectiveness, delivery, and organization of surgery and surgical services [7]. More recently, the NIHR Unit on Global Surgery was formed [10], to establish research hubs in low- and middle-income countries across the world. With the conduct of surgical trials growing in number, and becoming more geographically dispersed, ensuring that they are designed and analyzed appropriately is essential to support clinical decision making.

The assessment of surgical interventions is complex, because of the interacting components, such as the intervention itself, surgical expertise, and preoperative and postoperative care [11]. When designing randomized surgical trials, it is important to consider the potential existence and impact of surgical learning curves, where the surgical expertise increases throughout the course of the trial. Another important consideration is clustering. Clustering occurs when patient outcomes within center, surgical team, or surgeon are more similar than those from patients treated by different centers, teams, or surgeons.

Recognition and management of learning curves and clustering within clinical trials is recommended, [12] and may have increased relevance within the surgical field,

dependent upon the interventions being investigated and their routine use [11—15].

It is important to consider the significance of these aspects at trial outset, to ensure that the resulting trial is conducted and analyzed with the highest possible rigor. However, main trial publications often do not report deliberations and justifications for selected approaches [16]. To overcome this limitation, we investigate a cohort of applications for randomized surgical trials funded by the NIHR. This review will determine how learning and clustering, by center and surgeon, are managed at the design stage and accounted for in the intended analysis, and provide insight into who drives the decision making for these: the funder, guided by reviewers and panel members, or the researcher. We aim to provide a more detailed insight into current practice with regards to planning for, and acknowledging, the presence of learning and clustering when designing randomized surgical trials.

2. Materials and methods

2.1. Included studies

We sought to examine trials that had received funding from the NIHR from two funding streams, the Health Technology Assessment (HTA) program [17] and Efficacy and Mechanism Evaluation (EME) [18] program, in the United Kingdom, from 2012 to 2016. Research projects funded by these programs are either in response to a commissioning brief or an open investigator-led call. These funding streams were chosen as they are known to endorse high-quality research and were actively funding surgical research during this time [7]. An initial unpublished search indicated that this period would provide a reasonable cohort size to establish current practice. All randomized trials where the patient pathway involves a surgical intervention of any kind were eligible for inclusion.

2.2. Documents for review

The NIHR HTA and NIHR EME funding process involves a two-stage, peer-reviewed application process. Protocols and the commissioning brief (where applicable) were obtained from the open-access NIHR Journals Library [19]. The NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC) provided documentation not publicly available: project descriptions and applicant responses to reviewer comments.

2.3. Data extraction

A previously developed extraction form [16] was adapted for use on this cohort by E.J.C. and C.G. and approved by G.B., J.M.B., and J.A.C. (see Supplement A1). The extraction form was piloted on five applications initially and, as no further amendments were required, subsequently

used on all applications by a single assessor (E.J.C.). Data extracted were quality-checked through double data extraction by a second reviewer (A.R.H.) on 10% of all applications. A discrepancy rate was specified a priori such that if greater than 5% across all fields then a further 10% would be checked until the rate was below 5%. Discrepancies were jointly reviewed and agreement was reached; if agreement could not be reached, then a third reviewer (C.G.) was consulted.

Details on trial design, randomization stratification, sample size adjustment, predetermined center and surgeon credentials, outcomes, and planned statistical analyses that adjusted for center and surgeon were collected.

2.4. Statistical analysis

Quantitative items were summarized using descriptive statistics; no formal statistical comparisons were undertaken. Data were analyzed using SAS 9.3 (SAS Institute Inc., Cary, NC, USA). Open textual data items were categorized using NVivo qualitative data analysis software (QSR International Pty Ltd., Doncaster, Victoria, Australia, version 10, 2012). A confidentiality agreement with the NIHR Evaluation, Trials and Studies Coordinating Centre was signed prior to receiving the documentation. The raw data cannot therefore be made publicly available, and text extracts have been anonymized by removal of treatment or condition identifiers. Deleted text is denoted by *[…]* and the addition of words or replaced words is denoted by *[words]* to aid understanding.

3. Results

3.1. Cohort details

The NETSCC compiled a report listing all surgery randomized controlled trials funded by the NIHR HTA and NIHR EME funding streams within the eligible period. Sixty potentially eligible studies were identified, of which 49 (82%) met the eligibility criteria following further central screening (Fig. A1).

3.2. Double data extraction

Five articles were randomly selected from the eligible studies for double data extraction. Of 155 variables checked, two discrepancies were identified (1% error rate).

3.3. Cohort summary

The majority of the applications were funded by the NIHR HTA ($n = 44/49$, 89%) and had start dates from 2014 onward ($n = 37/49$, 76%) (see Table 1).

Documents for review consisted of commissioning briefs ($n = 15/49$, 31%), project descriptions ($n = 40/49$, 82%), applicant responses to board and peer review comments ($n = 40/49$, 82%) and protocols ($n = 42/49$, 86%).

Either the protocol or project description was available for all applications (see Table 1).

One application consisted of two distinct RCTs, herein treated as two separate trials.

3.4. Trial demographics

Trials were primarily two-armed ($n = 45/50$, 90%) and of a parallel design ($n = 49/50$, 98%). Eight did not use a pilot or feasibility study ($n = 8/50$, 16%) [20]. In 11 studies ($n = 11/50$, 22%), surgery was not the intervention of interest and delivered as part of the patient pathway. Where surgery was the intervention of interest ($n = 39/50$, 78%), 21 compared against surgery, for example, minimal access vs. open surgery ($n = 21/39$, 54%). The remaining 18 compared surgery against a nonsurgical comparator (medical comparator, e.g., injection vs. surgery: $n = 7/39$, other e.g., active monitoring and surgery vs. active monitoring only: $n = 11/39$) (see Tables A1 and 2).

3.5. Recruitment and randomization

Patients were the randomization unit in all trials and were primarily allocated to equal groups ($n = 48/50$, 96%). The majority stratified randomization ($n = 46/50$, 92%). In trials comparing two surgeries, there were no expertise-based designs [21]. Table A2 provides more detail.

Almost all studies were multicentere ($n = 49/50$, 98%), with over half stratifying by center ($n = 28/49$, 57%). Of the 21 that did not stratify by center, one provided justification for not stratifying by centre and this was related to concern over allocation concealment:

> *"To reduce the risk of the randomization sequence being predictable we will not stratify by center, which in addition to using randomly selected permuted blocks, will make the allocation sequence unpredictable for individual trial centers."*

Twenty-two trials had multiple surgeons within each center, of which eight stratified the randomization accordingly ($n = 8/22$, 36%). Two surgeon-stratified trials followed funder recommendation.

> *"We have made a number of changes since the first application…randomization will be stratified according to [stratification 1], [stratification 2], and according to consultant surgeon."*

In trials reported as multicenter and multisurgeon ($n = 21$), two stratified for both center and surgeon, eleven center only, six surgeon only, and two stratified for neither.

Three trials were international, of which one stratified randomization on randomized within a UK, or non-UK, center.

Table 2 provides more detail.

Table 1. Cohort summary

Item	Category	n	N	n/N%
Number of RCTs in application	One	48	49	98%
	Two	1	49	2%
Funder	NIHR HTA	44	49	90%
	NIHR EME	5	49	10%
Lead institution region	East	1	49	2%
	East Midlands	4	49	8%
	London	10	49	20%
	North East	7	49	14%
	North West	2	49	4%
	Scotland	10	49	20%
	South East	3	49	6%
	South West	4	49	8%
	Wales	2	49	4%
	West Midlands	4	49	8%
	Yorkshire and the Humber	2	49	4%
Trial start year	2012	3	49	6%
	2013	9	49	18%
	2014	26	49	53%
	2015	3	49	6%
	2016	1	49	2%
	2017	7	49	14%
Source documents available[a]	Commissioning brief	15	49	31%
	Project description	40	49	82%
	Responses to board and peer review comments	40	49	82%
	Protocol	42	49	86%

Abbreviations: EME, Efficacy and Mechanism Evaluation; HTA, Health Technology Assessment; NIHR, National Institute for Health Research; RCTs, randomized controlled trials.

[a] Documents available: All applications with project description also had responses to board and peer review comments (n = 40). A minimum of either the protocol or the project description and responses to board and peer review comments were available for all applications.

3.6. Surgeon and center credentials

Center and surgeon credentials, or inclusion criteria of those delivering the intervention, were provided in 41 (n = 41/50, 82%) and 36 (n = 36/50, 72%) trials respectively (Table 3). Most common center credentials were case volume (n = 20) and required fields of expertise within center (n = 13). Examples of surgeon credentials were

Table 2. Stratification factors in multicenter and multisurgeon trials by intervention type

Nature of surgery delivered		Comparator	Number of trials in cohort	Stratified by center					Stratified by surgeon				
				Multicenter	Yes		No		Multisurgeon	Yes		No	
				N	n	n/N%	n	n/N%	N	n	n/N%	n	n/N%
As an intervention	Surgery	Alternative surgical procedure	13	13	5	38%	8	62%	6	4	67%	2	33%
		Change to a component of the same procedure	6	5	4	80%	1	20%	6	3	50%	3	50%
		Same procedure delivered at a different time point	2	2	1	50%	1	50%	0	0	—	0	—
	Medical		7	7	5	71%	2	29%	2	0	—	2	100%
	Other		11	11	5	45%	6	55%	3	0	—	3	100%
As part of patient pathway			11	11	8	73%	3	27%	5	1	20%	4	80%

Table 3. Center and surgeon credentials

Center level		Surgeon level	
Center credential provided	41	Surgeon credentials provided	36
Case volume	20 (48%)	Level of job role	16 (44%)
Fields of expertise within center	13 (32%)	Study-specific training	13 (36%)
Experience required without definition	9 (22%)	Experience required without definition	8 (22%)
Experience required with definition	8 (20%)	Oversight of supervision	7 (19%)
Good recruiting reputation	8 (20%)	Prior number of cases	7 (19%)
Experience required with definition	8 (20%)	Self-assessed ability	7 (19%)
Access to equipment required	7 (17%)	Equipoise	4 (11%)
Center to undertake trial-specific training	2 (5%)	Known to be good recruiters	3 (8%)
Demonstrated ability to participate	1 (2%)	Case volume	2 (6%)
Interest expressed in specific treatment	1 (2%)	Local practice relevant	1 (3%)
Prior number of cases required	1 (2%)		
Center delivers one treatment only	1 (2%)		

grade or experience ($n = 16$) and study-specific training ($n = 13$).

3.7. Trial outcomes related to learning and clustering

Forty-one applications explored outcomes that may reflect variability in center or surgeon skill (82%, Table 4). Common outcomes were safety events ($n = 36$); recovery from surgery ($n = 13$), and operative time ($n = 6$).

Surgeon level outcomes were experience of surgeons in trial, established through qualitative methods ($n = 3$); surgeon accuracy as a main trial outcome ($n = 1$); and expertise ($n = 1$), more specifically:

> "The first [feasibility] phase will establish [words] and a measure of surgical expertise."

3.8. Statistical considerations

3.8.1. Sample size calculation

There were no examples of sample size adjustment for clustering at a center level. Three applications adjusted

Table 4. Outcomes

Outcome	
Relevant outcome reported	41
Safety measures	36 (88%)
Recovery from surgery time	13 (32%)
Operative time	6 (15%)
Patient satisfaction with surgery	5 (12%)
Infection	4 (10%)
Experience of surgeons in trial[a]	3 (7%)
Surgeon accuracy	1 (2%)
Surgeon expertise[b]	1 (2%)

[a] Established using qualitative methods.
[b] Feasibility outcome.

the sample size for surgeon using an intraclass correlation coefficient and a fourth chose not to adjust although provided justification:

> "As this study is not evaluating surgery per-se, surgical experience is not a criterion for participation (all participants will be under the care of a consultant surgeon). In the context of [this] study, clustering by surgeon is not relevant to the sample size and can be ignored (on the basis that the intraclass correction is negligible)."

3.8.2. Exploratory analysis

Eight applications planned exploratory analysis considering differences by center. Three analyzed using descriptive statistics and three via a subgroup analysis: the first conducting a trial center-by-treatment effect analysis, the second comparing outcomes between more and less experienced centers, and the third exploring trends within centers over time. A sensitivity analysis adjusting for center effects was planned in one application. Learning within center was described in another.

> "The effect of experience in [comparator intervention] at each recruitment center will be studied to characterise the effect of the learning curve on clinical effectiveness, and also the effect on [standard intervention] outcomes."

Exploratory analyses considering differences by surgeon were planned in seven applications, of which three also explored by center. Two analyzed descriptively by surgeon grade and four via subgroup analysis: one modeled the learning curve using outcomes operation time and complications as a proxy to measure the task efficiency of the surgeon, one planned to explore trends and changes over time between experienced and less experienced surgeons, one via a qualitative analysis, and the final where patients were

Table 5. Planned statistical adjustments through analysis in multicenter and multisurgeon trials

	Center			Surgeon		
	n	*N*	*n/N%*	*n*	*N*	*n/N%*
Adjustment made	21	49	43%	15	22	68%
Approach to adjustment (type of effect)						
Fixed	0	21	–	0	15	–
Random	9	21	43%	6	15	40%
Time varying	0	21	–	0	15	–
Not specified	12	21	57%	9	15	60%
Randomization stratified by and adjustment made						
Yes	17	21	81%	9	15	60%

sampled for observations in theater according to their treating surgeons' grade. As with center, one application planned a sensitivity analysis that adjusted for surgeon.

3.8.3. Formal adjustment

Formal adjustment for multiple center or surgeon effect was planned in 21 and 15 applications, respectively. Table 5 provides more detail. When formally adjusting for center, nine planned to use a random effect and thirteen did not specify. Similarly, six planned to adjust for surgeon using a random effect and nine did not specify. Of the applications planning a formal adjustment, 17 ($n = 17/21$, 81%) of applications adjusting for center and nine ($n = 9/15$, 60%) adjusting for surgeon did so in addition to stratifying randomization by these variables.

The two applications that planned to stratify by both center and surgeon (Table 3) also planned formally adjusting analysis by these factors.

3.9. Funder-led considerations

3.9.1. Commissioning briefs

Of the 15 commissioning briefs, one permitted single-center studies and one required a multicenter setting. No other brief gave guidance with respect to number of centers. Two briefs identified surgical learning considerations as an issue to address: the first indicating outcomes may be independent of surgeon grade and the second:

> "Proposals should account for the possibility of a learning curve affecting the outcomes of [surgery]."

3.9.2. Changes driven by funder

Response to referee comments were available for 40 studies ($n = 40/49$, 82%). Fourteen examples of change within 12 applications were identified. Funder concerns led to sample size adjustment for surgeon ($n = 3$); randomization balanced for surgeon ($n = 2$) and center ($n = 1$); and improved generalizability by increasing the number of centers ($n = 3$):

> "The Board suggested that the team should consider the addition of a second center to demonstrate generalisability and help with recruitment."

In one application, funders requested applicants increase homogeneity in treatments and the applicants argued against this.

> "To ensure homogeneity in treatments we have consulted with our participating surgeons [and] the National [...] Registry and agreed to specify the use of a CE marked [device...there are three main devices]. Surgical trials that specify a single type of [device] are notoriously difficult to conduct and we do not believe such a design could recruit surgeons, nor would the outputs be generalisable."

Further considerations with regard to surgeon credentials ($n = 3$) and the impact of surgeon equipoise on recruitment ($n = 1$) were also funder driven.

> "The sample size has been increased from a total of [n] patients to a total of [1.4n] to take into account clustering of surgeon as per the feedback from the first stage."

4. Conclusions

This review has investigated the decision making behind intended design and analysis of 50 randomized surgical trials funded by the NIHR EME and NIHR HTA programs from 2012 to 2016. These results show frequent consideration of centers and surgeon impact during design, and these may be funder led, due to concerns around homogeneity or generalizability of results. This review provides a cross-sectional insight into current practice of researchers, and expectations of reviewers and funders, during trial design within two streams of a major UK funder [17,18].

The need for transparency around learning curves and clustering is highlighted within reporting of nonpharmacological interventions guidelines [22,23], and a review of the published literature identified a deficiency in adherence to these [16]. By contrast, this review identifies that

considerations to manage learning and clustering are made, by both researchers and funders, during development of trials funded by a prestigious body. For example, 30% of multicenter and 12% of multisurgeon studies reported a statistical adjustment of these within published manuscripts. This was 43% and 68%, respectively, in this cohort. When randomization was stratified by center or surgeon, this was accounted for in the analysis in 30% of multicenter and 40% of multisurgeon trials in the published manuscripts, as opposed to 81% and 60% in this cohort. In drawing this comparison, it is important to differentiate between the intended audiences. The detail required for a funding application, assessed by clinicians and methodologists/statisticians, may exceed that required to communicate results to a clinical audience. This demonstrates benefit in exploring unpublished trial documentation to understand approaches to trial design and analysis and highlights the need for improvements to transparent reporting.

The cohort included successful applications to the NIHR 2012 call for *Applied Health in Surgery* [7]. This call recognized the need to increase research-based evidence in surgery. Applications that evaluated technology-driven implanted or implantable medical devices, surgical procedures, or surgical services were invited. As a clinical trial is typically a major financial investment [24], applicants need to assure funders that their proposal is important, is well designed, and demonstrates scientific value to add to the current evidence base. Each application undergoes a peer review process, where ''experts'' critically review the trial to ensure standards are met in terms of design, quality, feasibility, acceptability, and importance of the topic [17,18]. A strength of this review is the insight into the designs proposed to funders, and impact of feedback on subsequently funded studies.

While the degree of learning and clustering will vary trial-to-trial, many interventions require surgical skill in their delivery regardless of whether the surgery is the intervention of interest. The impact of any potential imbalance in delivery on comparing interventions should be considered at trial outset routinely. Early and careful consideration will ensure that procedures are standardized as completely as possible such that, in severe cases, the trial team can alleviate any doubts about homogeneity raised by the medical community should the trial results be questioned [12]. These results indicate funder awareness of this early consideration, with one of the two examples of balancing randomization surgeon following recommendation being in a trial where surgery was not the intervention of interest.

When interpreting these results, it is important to consider the limitations of this review. First, only successful applications could be included because of confidentiality constraints. It is therefore not possible to determine whether the management of learning and clustering contributes to the success of the application. However, given that the application process consists of iterations whereby peer reviewers are able to request that researchers address paucities in their application, it is unlikely that a promising application, lacking in the appropriate considerations, would be deemed unsuitable for funding outright. More likely, researchers would be given the opportunity to make these considerations during this iteration process. Second, as part of this iterative review, it is possible that additional discussions at the funder board meetings did not make it in to the comments fed back to applicants. This could mean that funders raised these issues more frequently than this review suggests. Third, due to the nature of the grant application process, the funder impact observed may be in part due to an increased awareness of the reviewers involved. Fourth, this work has focused on a single funding body that primarily supports UK-based research. However, trials supported span a wide range of surgical specialties and health care conditions and results from this review will be generalizable to other funding bodies with a similar peer review process.

Fundamental to trial design and analysis is understanding the objectives. While considerations relating to clustering and learning effects are not widely reported in main trial publications, these results indicate both funders and researchers consider these aspects to address a specific research question. Such issues may have varying relevance depending on the overall design of the trial. A very pragmatic study may deliberately include surgeons and centers of all types and have less emphasis on expertise and learning, whereas the delivery of the intervention in more explanatory studies is critical and requires consideration during design and analysis. Another approach to overcoming these issues is to provide quality assurance of the intervention. Early work to develop methods to achieve this have been developed, and it is expected that this will expand in the future [25]. Furthermore, these results provide insight into the promising role of the funder as a driver to improving the, long criticized, surgical evidence base. The funder, who has influence over whether and how studies are carried out and has been suggested as a driver for improving the quality of research during the period of growth for surgical trials [3], can play a valuable role in ensuring that future trials do not have the same shortfalls as those in the past.

CRediT authorship contribution statement

Elizabeth J. Conroy: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Anna Rosala-Hallas:** Validation, Writing - review & editing. **Jane M. Blazeby:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Girvan Burnside:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Jonathan A. Cook:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Carrol Gamble:**

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.05.007.

References

[1] Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. J Clin Nurs 2002;12:77—84.

[2] Horton R. Surgical research or comic opera: questions, but few answers. Lancet 1996;13:347.

[3] McCulloch P, Feinberg J, Philippou Y, Kolias A, Kehoe S, Lancaster G, et al. Progress in clinical research and IDEAL. Lancet 2018;392:88—94.

[4] Ahmed AU, van der Sluis PC, Issa Y, Habaga IA, Gooszen HG, Flum DR, et al. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. Ann Surg 2013; 258(2):199—207.

[5] Masters J, Costa M. How to build a randomised controlled trial. And how to decide when it is appropriate. R Coll Surg Bull 2017;99(6):227—30.

[6] Blencowe N, Cook JA, Pinkney T, Rogers C, Reeves BC, Blazeby JM. Delivering successful randomized controlled trials in surgery: methods to optimize collaboration and study design. Clin Trials 2017;14:211—8.

[7] Applied Health Research in Surgery. National Institute for Health Research. 2012. Available at https://www.nihr.ac.uk/funding-and-support/documents/themed-calls/Surgery.pdf. Accessed January 9, 2019.

[8] Meara JG, Leather JM, Hagander L, Alkire BC, Alonso N, Ameh EA, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. Lancet 2015;386:569—624.

[9] Hu Y, Edwards BL, Brooks KD, Newhook T, Slingluff CL Jr. Recent trends in national institute of health funding for surgery: 2003 to 2013. Am J Surg 2015;209:1083—9.

[10] National Institute for Health Research Unit on Global Surgery. Global Surg. 2017. Available at http://globalsurg.org/about/. Accessed January 9, 2019.

[11] Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, et al. Challenges in evaluating surgical innovation. Lancet 2009;374: 1097—104.

[12] ICH E9 Expert Working Group. Statistical principles for clinical trials: ICH harmonized tripartite guideline. Stats Med 1999;18: 1905—42.

[13] Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. Clin Trials 2004;1:421—7.

[14] Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. Clin Trials 2005;2:152—62.

[15] Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. Trials 2009;10:9.

[16] Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, Gamble C. Randomized trials involving surgery did not routinely report considerations of learning and clustering effects. J Clin Epidemiol 2018;107:27—35.

[17] Health Technology Assessment. National Institute for Health Research. Available at https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technology-assessment/. Accessed January 9, 2019.

[18] Efficacy and Mechanism Evaluation. National Institute for Health Research. Available at https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/efficacy-and-mechanism-evaluation/. Accessed January 9, 2019.

[19] National Institute for Health Research Journals Library. National Institute for Health Research. Available at https://www.journalslibrary.nihr.ac.uk/journals/. Accessed March 8, 2019.

[20] Feasibility and Pilot Studies. National Institute for Health Research. Available at https://www.nihr.ac.uk/funding-and-support/documents/funding-for-research-studies/research-programmes/PGfAR/CCF-PGfAR-Feasibility-and-Pilot-studies.pdf. Accessed January 9, 2019.

[21] Deveraux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomise controlled trials. BMJ 2005;330:88.

[22] Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, for the CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008;148:295—309.

[23] Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, for the CONSORT Group. CONSORT statement for randomized trials for nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann Intern Med 2017;167:40—7.

[24] Clinical Trials Toolkit: funding proposal [Clinical Trials Toolkit web site]. Available at http://www.ct-toolkit.ac.uk/routemap/funding-proposal/. Accessed January 9, 2019.

[25] Blencowe NS, Mills N, Cook JA, Donovan JL, Rogers CA, Whiting P, et al. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. Br J Surg 2016; 103(10):1377—84.