Journal of Clinical Epidemiology 107 (2019) 27−35

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Randomized trials involving surgery did not routinely report considerations of learning and clustering effects

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

[a]Department of Biostatistics, University of Liverpool, 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

Accepted 5 November 2018; Published online 13 November 2018

Abstract

Objectives: To establish current practice of the management of learning and clustering effects, by treating center and surgeon, in the design and analysis of randomized surgical trials.

Study Design and Setting: The need for more surgical randomized trials is well recognized, and in recent years conduct has grown. Rigorous design, conduct, and analyses of such studies is important. Two methodological challenges are clustering effects, by center or surgeon, and surgical learning on trial outcomes. Sixteen leading journals were searched for randomized trials published within a two-year period. Data were extracted on considerations for learning and clustering effects.

Results: A total of 247 eligible studies were identified. Trials accounted for learning with 2% using an expertise-based design and 39% accounting for expertise by predefining surgeon credentials. One study analyzed learning. Clustering, by site and surgeon, was commonly managed by stratifying randomization, although one-third of center and 40% of surgeon stratified trials did not also adjust analysis.

Conclusion: Considerations for surgical learning and clustering effects are often unclear. Methods are varied and demonstrate poor adherence to established reporting guidelines. It is recommended that researchers consider these issues on a trial-by-trial basis, and report methods or justify where not needed to inform interpretation of results.

Keywords: Randomized controlled trial; Surgery; Clustering; Learning curve; Statistics

1. Introduction

The need for more surgical randomized trials is well recognized [1], and in recent years, the number of surgical trials has grown [2]. Further growth is expected with leading research organizations setting out to improve the evidence base on a global scale [3−7]. Ensuring that these trials are designed, conducted, and analyzed with the highest possible rigor will support clinical decision-making.

Interacting components in surgical interventions, such as the procedure itself, surgeon expertise, and aftercare, increase the complexity in assessing these interventions [8]. The existence and impact of surgical learning curves, where a surgeon's expertise increases throughout the course of a trial, should be considered. This may be particularly relevant when a trial is evaluating new interventions. Another important consideration is clustering, where variation in outcomes may be smaller between patients treated by the same surgeon or center than patients treated by different surgeons or centers. It is recommended that these issues are considered in multicenter clinical trials [9] and may have increased relevance within some surgical trials dependent on the interventions under investigation and their levels of routine use [8−12]. Recognizing and managing these components appropriately can therefore be a challenge when designing and analyzing such trials.

Communicating these considerations through complete and transparent reporting can aid appraisal and interpretation by the wider surgical community. Reporting standards for trials of nonpharmacologic treatments, such as surgery, have been established [13,14]. Among their requirements, reporting of items specifically relating to learning curves and clustering are recommended [13,14].

* 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).

https://doi.org/10.1016/j.jclinepi.2018.11.004

What is new?

Key findings

- A novel assessment of the management of surgical learning and clustering is presented.
- A lack of consideration for surgical learning and clustering is identified.

What this adds to what was known?

- This review represents practice across a wide variety of trials, both by surgical discipline and by geographic location, published within a cohort of leading surgical and medical journals.
- This review is timely as represents publications at a time of remarkable growth within the surgical field.

What is the implication and what should change now?

- When considerations are made, methods are varied and demonstrate poor adherence to established reporting guidelines.
- Recommendations are provided about when and how to address surgical learning and clustering in the design, conduct and analysis of randomised surgical trials.

Learning curves and clustering have so far been investigated in isolation, often within specific fields and including studies of observational design [15,16]. The objective of this work was to provide an up-to-date and comprehensive overview of current practice in randomized surgical trials with regards to the management of surgical learning and clustering effects in design and analysis.

2. Materials and methods

2.1. Included studies

This work sought to review reports of randomized surgical trials within the wider surgical literature. Articles for inclusion in the cohort were identified by undertaking an electronic search using SCOPUS from a subset of journals. These journals were identified as the ten leading English-language general surgical journals (1−10, Box A1) [17] plus six general medical journals (11−16, Box A1). The rationale for selecting leading surgical and general medical journals was the assumption of endorsing high standards of reporting when publishing randomized controlled trials (RCTs). Primary reports of the results of RCTs evaluating a surgical intervention or a nonsurgical intervention, which required surgery to be administered, published from January 1, 2014, to the date the search was conducted (February 11, 2016), were eligible. Duplicate publications, secondary analyses, and interim reports of RCTs were excluded. All RCTs meeting these criteria were included in this cohort.

2.2. Data extraction

Selected journals were screened for RCTs that meet the eligibility criteria. Supplement A1 lists the search strategy for SCOPUS. E.J.C. screened articles to identify those eligible for selection. Because of the nature of the intervention of interest, full texts were screened to determine eligibility. When suitability was unclear, a second reviewer (C.G.) was consulted.

A data extraction form was developed by two authors (E.J.C. and C.G.), revised based on feedback from G.B., J.A.C., and J.M.B., and subsequently piloted on 30 articles before roll out to all articles (see Supplement A2). Data were extracted from all articles 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 the articles. An error rate was specified a priori such that if greater than 5% across all fields, then a further 10% would be checked until the error rate was below 5%. Data were extracted from all published materials (main trial report and, where applicable, supplementary material).

Data were extracted on generic trial design, for example, randomization details and statistical analyses related to clustering and learning at a center and surgeon level. Predetermined center and/or surgeon credentials and variables relating to surgical learning or clustering, either as a definitive outcome or as a variable of interest, such as duration of operation or the number of operations by surgical level were collected.

2.3. Statistical analysis

Quantitative items were summarized using descriptive statistics; no formal statistical comparisons were undertaken. SAS 9.3 (SAS Institute Inc., Cary, NC, USA) was used. Open textual responses were categorized using NVivo qualitative data analysis software (QSR International Pty Ltd. Version 10, 2012).

3. Results

3.1. Article details

The search identified 874 reports (398 in 2014; 446 in 2015; and 30 in 2016 to date of extraction February 11, 2016), of which 247 were eligible. Figure A1 provides the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. Supplement A3 and Table A1 provide a list of eligible studies and summarize the cohort demographics respectively. Two surgical specialty journals and three general medical journals contributed most of the eligible articles (Table A1). When reported (n = 167/247, 68%), more than half of eligible trials were European funded (n = 92/167, 54%), and more than one-fourth were North American funded (n = 48/167, 29%).

Table 1. Trial design features and characteristics

Item	Category	*n*	*N*	*n*/*N* (%)
Trial type	Definitive study	240	247	97
	External pilot or feasibility [18]	7	247	3
Type	Cluster	3	247	1
	Cross-over[a]	2	247	1
	Factorial	6	247	2
	Parallel	231	247	94
	Sequential	5	247	2
Number of arms	2	224	247	91
	3	16	247	6
	4	6	247	2
	6	1	247	<1
Some trial personnel blinded	Yes	157	247	64
	No	47	247	19
	Not reported	43	247	17
Expertise design	Pure—professionals delivering only one intervention[b]	4	247	2
	Hybrid—some professionals could deliver both[c]	1	247	<1
Intervention of interest	Surgery occurred but was not intervention of interest	105	247	43
	Surgery occurred and was the intervention of interest	142	247	57
Comparator when surgery was intervention of interest[d]	Surgery	111	142	78
	Medical	10	142	7
	Other, such as active monitoring	25	142	18
Surgical comparison in trials comparing two surgeries	Comparing different components of the same intervention	68	111	61
	Different surgical interventions	38	111	34
	Different time points of the same intervention	5	111	5

[a] Includes designs in which each participant receives both interventions.

[b] Reason for design: trial exploring effects of different training techniques for surgeons ($n = 1$); surgeon equipoise and belief of potential impact of learning curve ($n = 1$); trial exploring delivery differences between two types of health professionals ($n = 1$); not provided ($n = 1$).

[c] Reason for design: surgical preference ($n = 1$).

[d] Four studies classified twice as three arm and, therefore, two comparators.

Twenty-five articles were randomly selected from the eligible articles for double data extraction. Of 1,025 variables checked, 12 errors were identified (1.2%).

3.2. *Trial rationale and design*

Design features and characteristics of the trials are summarized in Table 1. Included trials were typically of a parallel ($n = 231$, 94%), two-armed ($n = 224$, 91%) design. Sixty-four percent described approaches to blinding trial personnel ($n = 157$, 64%).

Within the cohort, more than half of the trials were reported as multicenter ($n = 130$, 53%) and two-thirds were reported as multiple surgeons/care providers ($n = 162$, 66%; Table 2).

Very few trials used an expertise-based design [19], where the health professionals deliver only one of the comparators ($n = 5$, 2%). One of these used a hybrid design where some health professionals could deliver both interventions. Care providers/surgeons were allocated to arm based on preference ($n = 2$, of which one was a hybrid), randomization ($n = 2$), and the research question ($n = 1$; Table 1).

3.3. *Intervention of interest*

Surgery occurred and was the intervention under evaluation in approximately 60% of trials ($n = 142$; Table 1). Three-quarters ($n = 111$, 78%) of these also had a surgical comparator. The majority of which compared different

Table 2. Randomization considerations

Item	Category	*n*	*N*	*n*/*N* (%)
Multiple or single center trial	Multiple	130	247	53
	Single	101	247	41
	Not reported	16	247	6
Multiple or single care provider trial	Multiple	162	247	66
	Single	22	247	9
	Not reported	63	247	25
Randomization stratified	Yes	123	247	50
	No	124	247	50
If yes, randomization stratified by	Center and care provider	2	123	2
	Center	77	123	63
	Care provider	8	123	6
	Neither	36	123	29
Allocation of care provider	Pure—professionals delivering only one intervention			
	Defined by research question	1	4	25
	Preference	1	4	25
	Randomized	2	4	50
	Hybrid—some professionals could deliver both			
	Preference	1	1	100

components of the same intervention (n = 68/111, 61%), one-third compared different surgical interventions (n = 38/111, 34%), and a small number compared different time points of the same intervention, such as early or delayed surgery (n = 5/111, 5%). Box A2 gives an example from each category where surgery was a comparator. In trials where the intervention under evaluation was not surgery, surgery was a cointervention (n = 105, 43%; Table 1), for example, a trial of neoadjuvant chemotherapy and surgery where surgery was the same in both arms.

3.4. Center and surgeon credentials

Predefined center and surgeon credentials were reported in 41% of trials (n = 101; Table 3). This included 95 of the 162 multi surgeon trials, with common definitions being a set prior number of cases (n = 27) or a specific level or job role, for example consultant (n = 22). Fourteen trials reported criteria at center level, of which nine reported these alongside surgeon criteria. Examples of reported criteria are summarized in Table 3.

Table 3. Center and surgeon credentials

Item	Category	*n*	*N*	*n*/*N* (%)
Credentials defined	No, or not reported	146	247	59
	Yes	101	247	41
If credentials defined, at what level	Both center and surgeon	9	101	9
	Center only	5	101	5
	Surgeon only	87	101	87
Center credentials	Experience required without definition	8	14	57
	Prior number of cases defined	5	14	36
	Piloted technique	1	14	7
	Study-specific training	1	14	7
Surgeon credentials	Experience required without definition	29	96	30
	Prior number of cases	27	96	28
	Level or job role	22	96	23
	Study-specific training	21	96	22
	Oversight or supervision	17	96	18
	Local practice followed	10	96	10
	Experience in years	3	96	3
	Quality control by video	2	96	2

Table 4. Adjustment and stratification by comparator in multi surgeon trials

		Stratified by surgeon				Analysis adjusts for surgeon			
		Yes		No		Yes		No	
Comparator of interest	*N*	*n*	*n*/*N* (%)	*n*	*n*/*N* (%)	*n*	*n*/*N* (%)	*n*	*n*/*N* (%)
Totals	162	10	6	152	94	20	12	152	88
Surgery occurred but not intervention of interest	66	2	3	64	97	4	6	62	94
Surgery vs. medical	8	0	0	8	100	1	12	7	88
Surgery vs. other	20	0	0	20	100	1	5	19	95
Surgery vs. surgery									
Different components of the same intervention	45	3	7	42	94	7	16	38	84
Different surgical interventions	23	5	22	18	78	8	35	15	65
Different time points of the same intervention	3	0	0	3	100	0	0	3	100

Three trials counted across two comparator types as three arm studies.

3.5. Randomization

Table 2 provides a summary of randomization considerations. Half stratified the randomization (n = 123, 50%), using methods such as block randomization or minimization. Seventy-nine of the 130 multicenter trials stratified by center (61%), and ten of the 162 multiple surgeon trials stratified by surgeon (n = 10/162, 6%). Of the surgeon stratified trials (n = 10), half were trials comparing different components of the same intervention (n = 5; Table 4). Two trials stratified by both center and care provider/surgeon, and almost half stratified by neither (n = 75/162, 46%). Table A2 provides further exploration of the stratification approach within multi surgeon trials.

3.6. Considerations of learning and clustering of centers and surgeons

Variables reported relating to learning (Box A3) were background or level of surgeon or center (n = 14, 5%, of which one gave both); experience in years (n = 5, 2%); number of operations by surgeon level (n = 13, 5%); or over time (n = 1, <1%). Operation time was most commonly reported (n = 82, 33%). Variables relating to clustering (Box A3) were number of patients by region (n = 1, <1%); center (n = 39, 16%); surgeon (n = 13, 5%); the number of surgeons per center (n = 1, <1%); and the overlap of surgeons between arms (n = 2, 1%).

Of the 79 multicenter trials that stratified by center, one-third (n = 25) reported within center descriptive data, for example caseload. Likewise, of the ten stratified multicare provider trials, half (n = 5) reported descriptive data (Table A3).

Outcomes potentially relevant to clustering or learning, for example length of operation, are presented in Box A4. Eighty percent (198/247) reported on at least one outcome relevant to clustering or learning curves, with the most commonly reported being safety events (n = 129, 51%) and infection (n = 46, 19%).

3.7. Analysis adjustment of centers and surgeons

Center or care provider, when used to stratify the randomization process, was used to adjust the analysis in one-third of trials (n = 26/79, 33%, Table 5). Of the ten trials that stratified by care provider, four made analysis adjustments. One-third of multicenter (n = 45, 35%) and almost 90% of multi care provider trials (n = 140, 86%) neither stratified randomization nor made analysis adjustments.

Table 5. Stratification of randomization by analysis adjustment by center and care provider

		Analysis stratified				
		Yes				
		Total	All outcomes	Primary outcome only	Secondary outcomes only	No
Stratification factor	*N*	*N* (%)	*n* (%)	*n* (%)	*n* (%)	*N* (%)
Center						
Yes	79	26 (33)	10 (38)	14 (54)	2 (8)	53 (67)
No	51	6 (12)	2 (33)	4 (66)	0 (0)	45 (88)
Care provider						
Yes	10	4 (40)	2 (50)	2 (50)	0 (0)	6 (60)
No	152	12 (8)	4 (33)	6 (50)	2 (17)	140 (92)

Table 6. Statistical adjustment for multiple center and surgeon effects in primary or secondary analyses

		Center			Surgeon		
Item	**Category**	**n**	***N***	***n*/*N* (%)**	***n***	***N***	***n*/*N* (%)**
Analyses to address the potential effect planned	Yes	39	130	30	20	162	12
	No, but considered	2	130	2	1	162	1
	No	89	130	68	142	162	88
If yes, approach used	Term in regression model	32	39	82	15	20	75
	Separate exploratory analysis	4	39	10	0	20	0
	Other approach	3	39	8	3	20	15
Effect type where term in regression model	Fixed	1	32	3	2	15	13
	Random	16	32	50	6	15	40
	Time varying	0	32	0	1	15	7
	Unclear	15	32	47	6	15	40

Thirty-nine trials considered center effect in analysis of primary or secondary outcomes (16%; Table 6). When reported, adjustment using a random effect was more common (n = 16) than fixed effects (n = 1). Adjustments were applied to all outcomes in one-third of trials and to primary outcome only in almost half. Other approaches included: a sensitivity analysis excluding the center with the largest number of participants and center being used as a predictor to impute missing values.

Twenty trials considered surgeon effect in analysis of primary or secondary outcomes (8%; Table 6), with two-fifths of these being trials comparing different surgical interventions (n = 8, 40%; Table 4). Adjustments were applied to all outcomes in one-third of trials and to primary outcome only in 40%. Other approaches were to explore safety of surgeons in delivering interventions in a separate paper and to consider “run in” patients where the first 100 patients were randomized separately in analysis.

4. Conclusions

This review examines methods for addressing learning and clustering effects within a large cohort of 247 randomized surgical trials. Most commonly, learning effects were addressed in the design of the trial by surgeon or center participation requirements, for example, number of previous operations. Expertise-based studies were rare, although some may have been expertise based in delivery but not reported as such [20]. One study conducted a formal investigation of the learning curve using a time-varying treatment effect. Clustering was also most commonly accounted for in the design stage by stratifying the randomization process by center and/or surgeon. However, in most cases, the analysis was not then adjusted to reflect this [21].

Numerous examples in the literature demonstrate the presence of a learning curve and investigate the impact on outcomes over time [22,23]. In the surgical field, the appropriateness of making considerations for surgeon in an individual trial should be considered against how commonplace and stabilized the procedure or intervention are within routine practice. For example, consideration may be given to whether the trial is comparing established practices, established practices with minor differences, or entirely different or radical new procedures. Formal analysis of surgical learning was rare. When triallists consider the learning curve to be of interest, for example early phase studies involving radical new procedures, established statistical methods that allow the learning profile to be explored may be considered [24].

Approaches to manage clustering at the surgeon level were less prevalent than at site level within this cohort. This may be appropriate reflecting on the nature of the interventions being compared and their routine use. Impact of care bundles, for example pre- and post-surgical care, may be considered to exert a greater influence on outcomes than individual surgeon. These aspects of care are typically center-driven effects. Furthermore, a large cohort analysis of cardiac patients determined that 95% of variation in the outcome of interest was explained by patient risk factors, with surgeon and center contributing only 2%–3% respectively [25]. This raises the question of the importance of adjusting for surgeon particularly where the volume of data available limits the extent of modeling techniques. It is important to note that when surgeon and/or site are prognostic indicators in a trial, the randomization of the trial will often be balanced for this, commonly through stratifying the randomization process [9,26,27]. However, the subsequent analysis should be adjusted for these chosen stratification factors. Failure to adjust following stratification can inflate p-values and confidence interval widths potentially creating erroneous conclusions of no treatment benefit [21]. Within this cohort, one-third of site stratified trials and 40% of surgeon stratified trials reported making necessary adjustments to the analysis.

This review has identified potential deficiencies in the design and analysis of surgical trials. The regulatory

Box 1. Considerations and recommendations for design and analysis

Learning curve considerations and recommendations by scenario

Scenario	**Recommendations**
Interventions delivered by the same specialty and/or surgeons and:	
Delivered routinely within clinical practice	LC-1
Delivered routinely within clinical practice, with one intervention being a minor modification of the other	LC-1, LC-2, LC-4
Radical new procedure being compared with intervention commonly used within routine practice	LC-1, LC-2, LC-4, LC-5
Interventions delivered by different specialties and/or surgeons and:	
Delivered routinely within clinical practice	LC-1, LC-2, LC-3
Radical new procedure being compared with intervention commonly used within routine practice	LC-1, LC-2, LC-3, LC-5

Recommendations to mitigate any potential learning effect

LC-1	Consider defining care provider experience required to deliver the interventions.
LC-2	Consider whether trial specified training, at site or surgeon level, is required.
LC-3	Consider the appropriateness of an expertise-based vs. conventional design.
LC-4	Consider monitoring of protocol adherence and treatment delivery
LC-5	Consider whether it is appropriate to explore surgical learning as a secondary analysis of interest.

Clustering considerations and recommendations by scenario

Scenario	**Recommendations**
Randomization has to be performed at center level for logistical, not prognostic, reasons.	C-1
Randomization has to be performed at surgeon level for logistical, not prognostic, reasons.	C-2
Care bundle (pre- and post-operative care) varies between center.	C-1, C-3, C-4, C-5, C-6
Treatment delivery within site, or surgeon, may differ because of routine practice.	C-4, C-5, C-6
Center is a known prognostic indicator of outcome, for example because of patient population.	C-1, C-3, C-7
Surgeon is considered a prognostic indicator of outcome	C-2, C-3, C-7

Recommendations to mitigate any potential effect

C-1	Consider balancing randomization, as appropriate, with respect to center through stratification or minimization.
C-2	Consider balancing randomization, as appropriate, with respect to surgeon through stratification or minimization.
C-3	Balancing randomization can introduce correlation in outcomes within strata, analysis should subsequently be adjusted for prognostic factors on which the randomization is based to avoid potentially inflated p values and loss of power.
C-4	Consider stricter protocol requirements for treatment delivery.
C-5	Consider increasing monitoring of protocol adherence and treatment delivery.
C-6	Consider treatment effects across centers, these should be explored routinely to appropriately consider the generalizability of results. Not that treatment differences observed may be because of factors irrelevant of these. In this case, exploratory analysis into other factors may be warranted.
C-7	Regardless of randomization balancing factors, consider adjusting the analysis for prognostic factors. Note, interpretation of unadjusted analysis may be impacted.

governance of surgical trials is not comparable to pharmaceutical trials, however many of the requirements are directly relevant [9,26–28]. The ICH E9 Statistical Principles for Clinical Trials document discusses the variable reasons for conducting multicenter trials and the importance of defining the center appropriately, either by center or investigator [9]. This is directly applicable to surgical trials. Further guidance states that the potential for differential treatment effects across centers should be explored, with individual center results being reported and treatment-by-center interactions considered in the absence of homogeneity [9,28]. Our results show that practice does not follow this guideline, with one-third of multicenter studies, and 13% of multi surgeon studies, reporting approaches to check for differential outcome effects or justifying not doing so. It is important to remember that heterogeneity may be caused by factors not related to the surgeon. Heterogeneity may be explained at the center level, for example by differences in patient demographics, or at the level of the care provider/surgeon, for example because of variation in case mix complexity. Existence of heterogeneity between centers has implications for generalizability of study results and should be routinely investigated to appropriately consider generalizability. There is an absence of guidelines focusing on learning curves, which may be due to them originating specifically for medicinal trials. Other reasons for lack of guidance may be due to: expectations that learning curves are not expected in trials comparing commonly used practices; expectations that learning is suitably addressed, in training and selection of care providers, before trial commencement; or due to difficulties in measuring surgeon expertise, with method often being imperfect and subject to other influences, such as case mix.

The need for transparency around learning curves and clustering are highlighted within the guidelines on reporting of nonpharmacologic interventions [13,14]. This review identified poor adherence to these reporting guidelines with key requirements missing or only partially reported. Coupled with the poor adherence to good statistical practice guidelines [9], limitations in reporting may strengthen the concerns by health professionals that surgical research is of a poor quality, as this can ultimately lead to ill founded clinical decisions [13,14].

When interpreting these results, it is important to consider the limitations of this review. First, this cohort was restricted to top surgical and medical journals; although advantageous as it provides a wide variety of trials by surgical discipline and geographic location, these trials are more likely of a higher quality and better methodological practice because of wider adoption of reporting guidelines [29]. Second, this cohort is cross-sectional and, therefore, does not consider changes over time. However, because of the recent growth in surgical trials, and the establishment of reporting guidelines for trials such as surgery, it is likely that little is to be gained from reviewing more dated literature. Finally, drawing conclusions based on published articles may be hindered by a lack of transparent reporting. Because of word count constraints and within journal requirements, authors may not have been able to fully report methods used despite all available supplementary material being searched during this review. Further insight into methods used could have been obtained by interviews with authors although this would be resource intensive. Further insight into current practice could be informed by contact with current surgical trialists and statisticians, or by exploring trial documentation that may not be published, such as grant applications or protocols.

Fundamental to the design and analysis of a trial is the understanding of the trial objectives. Many multicenter trials are multicenter not because of interests in how treatment effects vary by center or surgeon but because of logistical considerations, or to provide a better basis for the subsequent generalization of its findings and to ensure sufficient availability of the patient population. Considerations and recommendations for design and analysis are presented for surgical learning and clustering in Box 1, based on current guidelines and recommendations through example scenarios [9,13,14,21,26−28]. These aspects of trial design and analysis demonstrate the need for early and continued expert statistical input [16].

Acknowledgments

None.

This research was funded by the National Institute for Health Research Doctoral Fellowship Program (Ref: DRF-2015-08-082). E.J.C. was funded through this fellowship program. J.M.B. and J.A.C. were partly funded by the Medical Research Council ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures—MR/K025643/1); J.M.B. is also supported by the National Institute for Health Research Biomedical Research Center at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol.

The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2018.11.004.

References

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

[2] Ahmed Ali U, 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.

[3] 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). https://doi.org/10.1308/rcsbull.2017.227.

[4] Blencowe NS, 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.

[5] Applied Health Research in Surgery. [National Institute for health research web site] 2012. Available at https://www.nihr.ac.uk/funding-and-support/documents/themed-calls/Surgery.pdf. Accessed November 6, 2017.

[6] Meara JG, Leather AJ, 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.

[7] 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.

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

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

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

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

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

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

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

[15] Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. Health Technol Assess 2001;5:1−79.

[16] Oltean H, Gagnier JJ. Use of clustering analysis in randomized controlled trials in orthopaedic surgery. BMC Med Res Methodol 2015;15:17.

[17] Mahawar KK, Kumar G, Malviya A. Who publishes in leading general surgical journals? The divide between the developed and developing worlds. Asian J Surg 2006;29(3):140−4.

[18] Feasibility and pilot studies. National Institute for Health Research website. 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 September 21, 2018.

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

[20] Cook JA, Elders A, Boachie C, Bassinga T, Fraser C, Altman DG, et al. A systematic review of the use of an expertise-based randomised controlled design. Trials 2015;16:241.

[21] Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. BMJ 2012;345:e5840.

[22] Doğan S, Bayraktar C. Endoscopic tympanoplasty: learning curve for a surgeon already trained in microscopic tympanoplasty. Eur Arch Otorhinolaryngol 2017;274(4):1853−8.

[23] Wu W, Xu J, Wen W, Yu Y, Xu X, Zhu Q, et al. Learning curve of totally thoracoscopic pulmonary segmentectomy. Front Med 2018; 12(5):586−92.

[24] Papachristofi O, Jenkins D, Sharples L. Assessment of learning curves in complex surgical interventions: a consecutive case-series study. Trials 2016;17:266.

[25] Papachristofi O, Klein AA, Mackay J, Nashef S, Fletcher N, Sharples LD, Association of Cardiothoracic Anaesthesia and Critical Care (ACTACC). Effect of individual patient risk, centre, surgeon and anaesthetist on length of stay in hospital after cardiac surgery: Association of Cardiothoracic Anaesthesia and Critical Care (ACTACC) consecutive cases series study of 10 UK specialist centres. BMJ Open 2017;7:e016947.

[26] European Medicines Agency Science Medicines Health. Guideline on adjustment for baseline covariates in clinical trials. 2015. [European Medical Agency web site]. Available at https://www.ema.europa.eu/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf. Accessed October 5, 2018.

[27] ICH E6 harmonised tripartite guideline: guideline for good clinical practice E6 (R1). Step 4, 10 June 1996 [ICH harmonisation for better health web site]. Available at https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf. Accessed January 26, 2018.

[28] ICH E3 harmonised tripartite guideline: structure and content of clinical study reports E3. Step 4, 30 November 1995 [ICH harmonisation for better health web site]. Available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf. Accessed February 22, 2018.

[29] Shantikumar S, Wigley J, Hameed W, Handa A. A survey of instructions to authors in surgical journals on reporting by CONSORT and PRISMA. Ann R Coll Surg Engl 2012;94(7): 468−71.