

## TITLE PAGE

### **Funders improve the management of learning and clustering effects through design and analysis of randomised trials involving surgery**

Mrs Elizabeth J Conroy (EJC) MSc. [ejconroy@liv.ac.uk](mailto:ejconroy@liv.ac.uk) \* <sup>a</sup>

Mrs Anna Rosala-Hallas (ARH) MSc. [arosala@liverpool.ac.uk](mailto:arosala@liverpool.ac.uk) <sup>a</sup>

Prof. Jane M Blazeby (JMB) MD. [j.m.blazeby@bristol.ac.uk](mailto:j.m.blazeby@bristol.ac.uk) <sup>b</sup>

Dr. Girvan Burnside (GB) PhD. [gburnsid@liverpool.ac.uk](mailto:gburnsid@liverpool.ac.uk) <sup>c</sup>

Assoc Prof. Jonathan A Cook (JAC) PhD. [jonathan.cook@ndorms.ox.ac.uk](mailto:jonathan.cook@ndorms.ox.ac.uk) <sup>3</sup>

Prof. Carrol Gamble (CG) PhD. [carrolp@liverpool.ac.uk](mailto:carrolp@liverpool.ac.uk) <sup>a, d</sup>

<sup>a</sup> Department of Biostatistics, University of Liverpool, a member of Liverpool Health Partners, UK.

<sup>b</sup> Centre for Surgical Research, Population Health Sciences, University of Bristol, UK.

<sup>c</sup> Centre for Statistics in Medicine, University of Oxford, UK.

<sup>d</sup> North West Hub for Trials Methodology Research, University of Liverpool, UK.

#### **\*Details of corresponding author and person to be contacted for reprint requests**

*Name:* Elizabeth J Conroy

*Postal Address:* 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

*Tel:* +44 151 795 8791

*Fax:* +44 151 795 8770

*Email address:* [ejconroy@liverpool.ac.uk](mailto:ejconroy@liverpool.ac.uk)

## ABSTRACT

### Objective

To provide insight into current practice in planning for, and acknowledging, the presence of learning and clustering effects, by treating centre and surgeon, when developing randomised surgical trials.

### Study design and setting

Complexities associated with delivering surgical interventions, such as clustering effects, by centre 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 four 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 pre-defined centre and surgeon credentials was common. One planned exploratory analysis of learning within centre, and two within surgeon. Clustering, by site and surgeon, was often managed through stratifying randomisation, 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

~~Whilst underreported in main publications,~~ This review indicates that researchers do consider impact of learning and clustering, by centres and surgeon, during trial development. Furthermore, the funder is identified as a potential driver of considerations.

### Running head

Management learning curve and clustering effects in surgical trials

**Word count**

3000

**Key words**

Randomized controlled trials; Surgery; Clustering; Learning curve; Statistics

## Abbreviations

EME	Efficacy and Mechanism Evaluation
HTA	Health Technology Assessment
NIHR	National Institute of Health Research
NETSCC	National Institute of Health Research Evaluation, Trials and Studies Coordinating Centre
RCT	Randomised Controlled Trial
UK	United Kingdom

## MANUSCRIPT TEXT

### 1. INTRODUCTION

Randomised controlled trials (RCTs) are recognised as providing the highest level of evidence, second only to systematic reviews of such trials.[1] The need for surgical randomised trials is well recognised [2, 3], and this has led to a push for growth in recent years. [3, 4] Leading research organisations are supporting this growth through establishing a number of initiatives and research objectives, ultimately aiming to improve of the global surgical evidence base. [5-10] One such initiative, set up by the United Kingdom's (UK) leading publically funded health research body, the National Institute of Health Research (NIHR), aimed to increase the volume of high quality research, across surgical disciplines, on the effectiveness, delivery and organisation 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 analysed appropriately is essential to support clinical decision-making.

The assessment of surgical interventions is complex, due to the interacting components, such as the intervention itself, surgical expertise and pre and post-operative care. [11] When designing randomised 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 centre, surgical team or surgeon, are more similar than those from patients treated by different centres, 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 therefore to consider the significance of these aspects at trial outset, to ensure that the resulting trial is conducted and analysed with the highest possible rigour. 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 randomised surgical trials funded by the NIHR. This review will determine how learning and clustering by centre 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 at the design stage.

## 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) programme [17] and Efficacy and Mechanism Evaluation (EME) [18] programme, in the UK, from 2012 to 2016. Research projects funded by these programmes 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 randomised 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 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 publically 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 EJC and CG and approved by GB, JAC, and JMB, 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 (EJC). Data extracted were quality checked through double data extraction by a second reviewer (ARH) 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 reached, if agreement could not be reached then a third reviewer (CG) was consulted.

Details on trial design, randomisation stratification, sample size adjustment, pre-determined centre and surgeon credentials, outcomes, and planned statistical analyses that adjusted for centre and surgeon were collected.

#### *2.4. Statistical Analysis*

Quantitative items were summarised using descriptive statistics; no formal statistical comparisons were undertaken. Data was analysed using SAS 9.3; SAS Institute Inc., Cary, NC, USA. Open textual data items; were categorised using NVivo qualitative data analysis software (QSR International Pty Ltd. 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 anonymised 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 randomised controlled trials funded by the HTA and 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 (**Figure 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 (1.3% error rate).

#### 3.3. Cohort summary

The majority of the applications were funded by the HTA (n=44/49, 89%) and had start dates from 2014 onwards (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 separate trials.

Item	Category	n	N	n/N%
Number of RCTs in application	One	48	49	98%
	Two	1	49	2%
Funder	HTA	44	49	90%
	EME	5	49	10%
Lead institution region	East	1	49	2%
	East Midlands	4	49	8%

Item	Category	n	N	n/N%
	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 <sup>1</sup>	Commissioning brief	15	49	31%
	Project description	40	49	82%
	Responses to board and peer review comments	40	49	82%
	Protocol	42	49	86%

<sup>1</sup> 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.

**Table 1: Cohort summary**

### 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 eighteen compared surgery against a non-surgical 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 **Table A1, Table 2**).

### 3.5. Recruitment and randomisation

Patients were the randomisation unit in all trials and primarily allocated to equal groups (n=48/50, 96%). The majority stratified randomisation (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 multi-centre (n=49/50, 98%), with over half stratifying by centre (n=28/49, 57%). Of the 21 that did not stratify by centre, only one provided justification which related to concern over allocation concealment:

*“To reduce the risk of the randomisation sequence being predictable we will not stratify by centre, which in addition to using randomly selected permuted blocks, will make the allocation sequence unpredictable for individual trial centres.”*

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

*“We have made a number of changes since the first application...randomisation will be stratified according to [stratification 1], [stratification 2], and according to consultant surgeon.”*

In trials reported as multi-centre and multi-surgeon (n=21), two stratified for both centre and surgeon, eleven centre only, six surgeon only, and two stratified for neither.

Three trials were international, of which one stratified randomisation on randomised within a UK, or non UK, centre.

**Table 2** provides more detail.

Nature of surgery delivered	Comparator		Number of trials in cohort	Stratified by centre					Stratified by surgeon				
				Multi-centre	Yes		No		Multi-surgeon	Yes		No	
					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 2: Stratification factors in multi-centre and multi-surgeon trials by intervention type**

### 3.6. Surgeon and centre credentials

Centre 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 centre credentials were case volume (n=20) and required fields of expertise within centre (n=13). Examples of surgeon credentials were grade or experience (n=16) and study specific training (n=13).

<b>Centre level</b>		<b>Surgeon level</b>	
Centre credential provided	41	Surgeon credentials provided	36
Case volume	20 (48%)	Level of job role	16 (44%)
Fields of expertise within centre	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%)
Centre 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%)		
Centre delivers one treatment only	1 (2%)		

**Table 3: Centre and surgeon credentials**

### 3.7. Trial outcomes related to learning and clustering

Forty-one applications explored outcomes that may reflect variability in centre 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.”*

<b>Outcome</b>	
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 <sup>1</sup>	3 (7%)
Surgeon accuracy	1 (2%)
Surgeon expertise <sup>2</sup>	1 (2%)

<sup>1</sup> Established using qualitative methods; <sup>2</sup> Feasibility outcome

**Table 4: Outcomes**

### 3.8. Statistical Considerations

#### 3.8.1. Sample size calculation

There were No-no examples of sample size adjustment for clustering at a centre level ~~were identified~~.

Three applications adjusted the sample size for surgeon using an intra class correlation coefficient (ICC) 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 centre. Three analysed using descriptive statistics and three via a subgroup analysis: the first conducting a trial centre by treatment effect analysis, the second comparing outcomes between more and less experienced centres, and the third exploring trends within centres over time. A sensitivity analysis adjusting for centre effects was planned in one application. Learning within centre was described in another.

*“The effect of experience in [comparator intervention] at each recruitment centre 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 centre. Two analysed descriptively by surgeon grade and four via subgroup analysis: one modelled 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 sampled for observations in theatre according to their treating surgeons' grade. As with centre, one application planned a sensitivity analysis that adjusted for surgeon.

### 3.8.3. Formal adjustment

Formal adjustment for multiple centre or surgeon effect was planned in 21 and 15 applications, respectively. **Table 5** provides more detail. When formally adjusting for centre, 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 centre and nine (n=9/15, 60%) adjusting for surgeon did so in addition to stratifying randomisation by these variables.

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

		Centre			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%
Randomisation stratified by and adjustment made	Yes	17	21	81%	9	15	60%

**Table 5: Planned statistical ~~Formal statistical~~ adjustments through analysis ~~made~~ in multi-centre and multi-surgeon trials**

### 3.9. Funder led considerations

#### 3.9.1. Commissioning briefs

Of the fifteen commissioning briefs, one permitted single centre studies and one required a multi-centre setting. No other brief gave guidance with respect to number of centres. 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, 81.6%). Fourteen examples of change within twelve applications were identified. Funder concerns led to sample size adjustment for surgeon (n=3); randomisation balanced for surgeon (n=2) and centre (n=1); and improved generalisability by increasing the number of centres (n=3):

*“The Board suggested that the team should consider the addition of a second centre 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 randomised surgical trials funded by the NIHR EME and NIHR HTA programmes from 2012 to 2016. These results show frequent consideration of centres and surgeon impact during design, and these may be funder led, due to concerns around homogeneity or generalisability 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 are highlighted within reporting of non-pharmacological interventions guidelines, [22, 23] and a review of the published literature identified a deficiency in adherence to these [16]. In 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 multi-centre and 12% of multi-surgeon studies reported a statistical adjustment of these within published manuscripts. This was 423% and 698% respectively in this cohort. When randomisation was stratified by centre or surgeon, this was accounted for in the analysis in 30% of multi-centre and 40% of multi-surgeon trials in the published manuscripts, as oppose 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 recognised the need to increase research-based evidence in surgery. Applications were invited that evaluated technology-driven implanted or implantable medical devices, surgical procedures or surgical services. As a clinical trial is typically a major financial investment, [24] applicants need to assure funders that their proposal is important, 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.

Whilst the degree of learning and clustering will vary trial-to-trial, many interventions require surgical skill in their delivery regardless of whether or not the surgery is the intervention of interest. The impact of any potential imbalance in delivery on comparing the intervention comparing interventions should be considered at trial outset routinely. Early and careful consideration will ensure that procedures are standardised 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 randomisation 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 due to 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 focussed 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 generalisable 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 in order 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 centres 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 criticised, surgical evidence base. The funder, who has influence over whether or not 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.

1  
2  
3  
4  
5  
6 **What is new?**  
7

- 8 • This review investigates successful funding applications comprising a wide variety of trials,  
9 both by surgical discipline and by geographic location, by a leading UK funder.  
10
- 11 • This review is timely as it comprises applications rewarded following a call by this funder  
12 recognising a need for an increase in evidence based surgical research.  
13
- 14 • A novel assessment of the decision making behind intended design and analysis with respect  
15 to the management of surgical learning and clustering is presented. Results indicate that while  
16 these considerations are under reported in main trial publications, funders and researchers  
17 alike appear to be aware of the need to manage these aspects at the trial design stage.  
18
- 19 • Insight into the promising role of the funder as a driver to improving the, long criticised,  
20 surgical evidence base is provided.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59

## TITLE PAGE

### Funders improve the management of learning and clustering effects through design and analysis of randomised trials involving surgery

Mrs Elizabeth J Conroy (EJC) MSc. [ejconroy@liv.ac.uk](mailto:ejconroy@liv.ac.uk) \* <sup>a</sup>

Mrs Anna Rosala-Hallas (ARH) MSc. [arosala@liverpool.ac.uk](mailto:arosala@liverpool.ac.uk) <sup>a</sup>

Prof. Jane M Blazeby (JMB) MD. [j.m.blazeby@bristol.ac.uk](mailto:j.m.blazeby@bristol.ac.uk) <sup>b</sup>

Dr. Girvan Burnside (GB) PhD. [gburnsid@liverpool.ac.uk](mailto:gburnsid@liverpool.ac.uk) <sup>c</sup>

Assoc Prof. Jonathan A Cook (JAC) PhD. [jonathan.cook@ndorms.ox.ac.uk](mailto:jonathan.cook@ndorms.ox.ac.uk) <sup>3</sup>

Prof. Carrol Gamble (CG) PhD. [carrolp@liverpool.ac.uk](mailto:carrolp@liverpool.ac.uk) <sup>a, d</sup>

<sup>a</sup> Department of Biostatistics, University of Liverpool, a member of Liverpool Health Partners, UK.

<sup>b</sup> Centre for Surgical Research, Population Health Sciences, University of Bristol, UK.

<sup>c</sup> Centre for Statistics in Medicine, University of Oxford, UK.

<sup>d</sup> North West Hub for Trials Methodology Research, University of Liverpool, UK.

#### **\*Details of corresponding author and person to be contacted for reprint requests**

*Name:* Elizabeth J Conroy

*Postal Address:* 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

*Tel:* +44 151 795 8791

*Fax:* +44 151 795 8770

*Email address:* [ejconroy@liverpool.ac.uk](mailto:ejconroy@liverpool.ac.uk)



60  
61  
62 **ABSTRACT**  
63  
64

65 **Objective**  
66

67  
68 To provide insight into current practice in planning for, and acknowledging, the presence of learning  
69 and clustering effects, by treating centre and surgeon, when developing randomised surgical trials.  
70

71  
72 **Study design and setting**  
73

74  
75 Complexities associated with delivering surgical interventions, such as clustering effects, by centre or  
76 surgeon, and surgical learning, should be considered at trial design. Main trial publications within the  
77 wider literature under-report these considerations  
78

79  
80 Funded applications, within a four year period, from a leading UK funding body were searched. Data  
81 were extracted on considerations for learning and clustering effects and the driver, funder or  
82 applicant, behind these.  
83  
84  
85  
86  
87

88  
89 **Results**  
90

91  
92 Fifty trials were eligible. Managing learning through establishing pre-defined centre and surgeon  
93 credentials was common. One planned exploratory analysis of learning within centre, and two within  
94 surgeon. Clustering, by site and surgeon, was often managed through stratifying randomisation, with  
95 81% and 60% respectively also planning to subsequently adjust analysis. One-third of responses to  
96 referees contained funder led changes accounting for learning and/or clustering.  
97  
98  
99

100  
101  
102 **Conclusion**  
103

104  
105 This review indicates that researchers do consider impact of learning and clustering, by centres and  
106 surgeon, during trial development. Furthermore, the funder is identified as a potential driver of  
107 considerations.  
108  
109

110  
111 **Running head**  
112

113  
114 Management learning curve and clustering effects in surgical trials  
115  
116  
117  
118

119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177

**Word count**

3000

**Key words**

Randomized controlled trials; Surgery; Clustering; Learning curve; Statistics

178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236

## Abbreviations

EME	Efficacy and Mechanism Evaluation
HTA	Health Technology Assessment
NIHR	National Institute of Health Research
NETSCC	National Institute of Health Research Evaluation, Trials and Studies Coordinating Centre
RCT	Randomised Controlled Trial
UK	United Kingdom

## **1. INTRODUCTION**

Randomised controlled trials (RCTs) are recognised as providing the highest level of evidence, second only to systematic reviews of such trials.[1] The need for surgical randomised trials is well recognised [2, 3], and this has led to a push for growth in recent years. [3, 4] Leading research organisations are supporting this growth through establishing a number of initiatives and research objectives, ultimately aiming to improve of the global surgical evidence base. [5-10] One such initiative, set up by the United Kingdom's (UK) leading publically funded health research body, the National Institute of Health Research (NIHR), aimed to increase the volume of high quality research, across surgical disciplines, on the effectiveness, delivery and organisation 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 analysed appropriately is essential to support clinical decision-making.

The assessment of surgical interventions is complex, due to the interacting components, such as the intervention itself, surgical expertise and pre and post-operative care. [11] When designing randomised 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 centre, surgical team or surgeon, are more similar than those from patients treated by different centres, 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].

296  
297  
298 It is important therefore to consider the significance of these aspects at trial outset, to ensure that  
299  
300 the resulting trial is conducted and analysed with the highest possible rigour. However, main trial  
301  
302 publications often do not report deliberations and justifications for selected approaches. [16] To  
303  
304 overcome this limitation, we investigate a cohort of applications for randomised surgical trials funded  
305  
306 by the NIHR. This review will determine how learning and clustering by centre and surgeon are  
307  
308 managed at the design stage and accounted for in the intended analysis, and provide insight into who  
309  
310 drives the decision-making for these: the funder, guided by reviewers and panel members, or the  
311  
312 researcher. We aim to provide a more detailed insight into current practice with regards to planning  
313  
314 for, and acknowledging, the presence of learning and clustering at the design stage.  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354

355  
356  
357 **2. MATERIALS AND METHODS**  
358

359 *2.1. Included studies*  
360

361  
362 We sought to examine trials that had received funding from the NIHR from two funding streams, the  
363 Health Technology Assessment (HTA) programme [17] and Efficacy and Mechanism Evaluation (EME)  
364 [18] programme, in the UK, from 2012 to 2016. Research projects funded by these programmes are  
365  
366 either in response to a commissioning brief or an open investigator led call. These funding streams  
367  
368 were chosen as they are known to endorse high quality research and were actively funding surgical  
369  
370 research during this time [7]. An initial unpublished search indicated that this period would provide a  
371  
372 reasonable cohort size to establish current practice. All randomised trials where the patient pathway  
373  
374 involves a surgical intervention of any kind were eligible for inclusion.  
375  
376  
377

378  
379 *2.2. Documents for review*  
380

381  
382 The NIHR HTA and EME funding process involves a two stage, peer reviewed application process.  
383  
384 Protocols and the commissioning brief (where applicable) were obtained from the open access NIHR  
385  
386 Journals Library [19] The NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC) provided  
387  
388 documentation not publically available: project descriptions and applicant responses to reviewer  
389  
390 comments.  
391

392  
393 *2.3. Data extraction*  
394

395  
396 A previously developed extraction form [16] was adapted for use on this cohort by EJC and CG and  
397  
398 approved by GB, JAC, and JMB, see **Supplement A1**. The extraction form was piloted on five  
399  
400 applications initially and, as no further amendments were required, subsequently used on all  
401  
402 applications by a single assessor (EJC). Data extracted were quality checked through double data  
403  
404 extraction by a second reviewer (ARH) on 10% of all applications. A discrepancy rate was specified a  
405  
406 priori such that if greater than 5% across all fields then a further 10% would be checked until the rate  
407  
408  
409  
410  
411  
412  
413

414  
415  
416 was below 5%. Discrepancies were jointly reviewed and agreement reached, if agreement could not  
417  
418 be reached then a third reviewer (CG) was consulted.  
419

420  
421 Details on trial design, randomisation stratification, sample size adjustment, pre-determined centre  
422  
423 and surgeon credentials, outcomes, and planned statistical analyses that adjusted for centre and  
424  
425 surgeon were collected.  
426

#### 427 428 *2.4. Statistical Analysis* 429

430 Quantitative items were summarised using descriptive statistics; no formal statistical comparisons  
431  
432 were undertaken. Data was analysed using SAS 9.3; SAS Institute Inc., Cary, NC, USA. Open textual  
433  
434 data items; were categorised using NVivo qualitative data analysis software (QSR International Pty Ltd.  
435  
436 Version 10, 2012). A confidentiality agreement with the NIHR Evaluation, Trials and Studies  
437  
438 Coordinating Centre was signed prior to receiving the documentation. The raw data cannot therefore  
439  
440 be made publicly available and text extracts have been anonymised by removal of treatment or  
441  
442 condition identifiers. Deleted text is denoted by [...] and the addition of words or replaced words is  
443  
444 denoted by [words] to aid understanding.  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472

### 3. RESULTS

#### 3.1. Cohort details

The NETSCC compiled a report listing all surgery randomised controlled trials funded by the HTA and 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 (**Figure 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 (1.3% error rate).

#### 3.3. Cohort summary

The majority of the applications were funded by the HTA (n=44/49, 89%) and had start dates from 2014 onwards (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 separate trials.

Item	Category	n	N	n/N%
Number of RCTs in application	One	48	49	98%
	Two	1	49	2%
Funder	HTA	44	49	90%
	EME	5	49	10%
Lead institution region	East	1	49	2%
	East Midlands	4	49	8%



532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590

Item	Category	n	N	n/N%
	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 <sup>1</sup>	Commissioning brief	15	49	31%
	Project description	40	49	82%
	Responses to board and peer review comments	40	49	82%
	Protocol	42	49	86%

<sup>1</sup> 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.

**Table 1: Cohort summary**

591  
592  
593 **3.4. Trial demographics**  
594

595  
596 Trials were primarily two-armed (n=45/50, 90%) and of a parallel design (n=49/50, 98%). Eight did not  
597  
598 use a pilot or feasibility study (n=8/50, 16%) [20]. In 11 studies (n=11/50, 22%), surgery was not the  
599  
600 intervention of interest and delivered as part of the patient pathway. Where surgery was the  
601  
602 intervention of interest (n=39/50, 78%), 21 compared against surgery, for example minimal access vs.  
603  
604 open surgery (n=21/39, 54%). The remaining eighteen compared surgery against a non-surgical  
605  
606 comparator (medical comparator e.g. injection vs. surgery: n=7/39, other e.g. active monitoring and  
607  
608 surgery vs. active monitoring only: n=11/39) (see **Table A1, Table 2**).  
609

610  
611 **3.5. Recruitment and randomisation**  
612

613  
614 Patients were the randomisation unit in all trials and primarily allocated to equal groups (n=48/50,  
615  
616 96%). The majority stratified randomisation (n=46/50, 92%). In trials comparing two surgeries, there  
617  
618 were no expertise-based designs [21]. **Table A2** provides more detail.  
619

620  
621 Almost all studies were multi-centre (n=49/50, 98%), with over half stratifying by centre (n=28/49,  
622  
623 57%). Of the 21 that did not stratify by centre, only one provided justification which related to concern  
624  
625 over allocation concealment:  
626

627  
628 *“To reduce the risk of the randomisation sequence being predictable we will not stratify by centre,*  
629  
630 *which in addition to using randomly selected permuted blocks, will make the allocation sequence*  
631  
632 *unpredictable for individual trial centres.”*  
633

634  
635 Twenty-two trials had multiple surgeons within each centre, of which eight stratified the  
636  
637 randomisation accordingly (n=8/22, 36%). Two surgeon-stratified trials followed funder  
638  
639 recommendation.  
640

641  
642 *“We have made a number of changes since the first application...randomisation will be stratified*  
643  
644 *according to [stratification 1], [stratification 2], and according to consultant surgeon.”*  
645  
646  
647  
648  
649

650  
651  
652 In trials reported as multi-centre and multi-surgeon (n=21), two stratified for both centre and surgeon,  
653 eleven centre only, six surgeon only, and two stratified for neither.  
654  
655

656  
657 Three trials were international, of which one stratified randomisation on randomised within a UK, or  
658 non UK, centre.  
659  
660

661  
662 **Table 2** provides more detail.  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708

709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749

Nature of surgery delivered	Comparator		Number of trials in cohort	Stratified by centre					Stratified by surgeon				
				Multi-centre	Yes		No		Multi-surgeon	Yes		No	
					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 2: Stratification factors in multi-centre and multi-surgeon trials by intervention type**

750  
751  
752 **3.6. Surgeon and centre credentials**  
753  
754

755 Centre and surgeon credentials, or inclusion criteria of those delivering the intervention, were  
756 provided in 41 (n=41/50, 82%) and 36 (n=36/50, 72%) trials, respectively (**Table 3**). Most common  
757 centre credentials were case volume (n=20) and required fields of expertise within centre (n=13).  
758  
759 Examples of surgeon credentials were grade or experience (n=16) and study specific training (n=13).  
760  
761  
762  
763

<b>Centre level</b>		<b>Surgeon level</b>	
Centre credential provided	41	Surgeon credentials provided	36
Case volume	20 (48%)	Level of job role	16 (44%)
Fields of expertise within centre	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%)
Centre 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%)		
Centre delivers one treatment only	1 (2%)		

764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790 **Table 3: Centre and surgeon credentials**  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808

809  
810  
811 **3.7. Trial outcomes related to learning and clustering**  
812  
813

814 Forty-one applications explored outcomes that may reflect variability in centre or surgeon skill (82%,  
815

816 **Table 4).** Common outcomes were safety events (n=36); recovery from surgery (n=13) and operative  
817  
818 time (n=6).  
819

820  
821 Surgeon level outcomes were experience of surgeons in trial, established through qualitative methods  
822  
823 (n=3); surgeon accuracy as a main trial outcome (n=1); and expertise (n=1), more specifically:  
824

825 “The first [feasibility] phase will establish [words] and a measure of surgical expertise.”  
826  
827

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 <sup>1</sup>	3 (7%)
Surgeon accuracy	1 (2%)
Surgeon expertise <sup>2</sup>	1 (2%)

828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849 <sup>1</sup> Established using qualitative methods; <sup>2</sup> Feasibility outcome  
850

851 **Table 4: Outcomes**  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867

868  
869  
870 3.8. *Statistical Considerations*  
871

872 3.8.1. Sample size calculation  
873  
874

875 There were no examples of sample size adjustment for clustering at a centre level. Three applications  
876 adjusted the sample size for surgeon using an intra class correlation coefficient (ICC) and a fourth  
877 chose not to adjust although provided justification:  
878

881 *“As this study is not evaluating surgery per-se, surgical experience is not a criterion for participation*  
882 *(all participants will be under the care of a consultant surgeon). In the context of [this] study, clustering*  
883 *by surgeon is not relevant to the sample size and can be ignored (on the basis that the intraclass*  
884 *correction is negligible”.*  
885  
886  
887  
888  
889  
890  
891

892 3.8.2. Exploratory analysis  
893  
894

895 Eight applications planned exploratory analysis considering differences by centre. Three analysed  
896 using descriptive statistics and three via a subgroup analysis: the first conducting a trial centre by  
897 treatment effect analysis, the second comparing outcomes between more and less experienced  
898 centres, and the third exploring trends within centres over time. A sensitivity analysis adjusting for  
899 centre effects was planned in one application. Learning within centre was described in another.  
900

901 *“The effect of experience in [comparator intervention] at each recruitment centre will be studied to*  
902 *characterise the effect of the learning curve on clinical effectiveness, and also the effect on [standard*  
903 *intervention] outcomes.”*  
904  
905  
906  
907  
908  
909  
910

911 Exploratory analyses considering differences by surgeon were planned in seven applications, of which  
912 three also explored by centre. Two analysed descriptively by surgeon grade and four via subgroup  
913 analysis: one modelled the learning curve using outcomes operation time and complications as a proxy  
914 to measure the task efficiency of the surgeon, one planned to explore trends and changes over time  
915 between experienced and less experienced surgeons, one via a qualitative analysis and the final where  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926

927  
928  
929 patients were sampled for observations in theatre according to their treating surgeons' grade. As with  
930  
931 centre, one application planned a sensitivity analysis that adjusted for surgeon.  
932

### 933 934 3.8.3. Formal adjustment 935

936  
937 Formal adjustment for multiple centre or surgeon effect was planned in 21 and 15 applications,  
938  
939 respectively. **Table 5** provides more detail. When formally adjusting for centre, nine planned to use a  
940  
941 random effect and thirteen did not specify. Similarly, six planned to adjust for surgeon using a random  
942  
943 effect and nine did not specify. Of the applications planning a formal adjustment, 17 (n=17/21, 81%)  
944  
945 of applications adjusting for centre and nine (n=9/15, 60%) adjusting for surgeon did so in addition to  
946  
947 stratifying randomisation by these variables.  
948  
949

950 The two applications that planned to stratify by both centre and surgeon (Table 3), also planned  
951  
952 formally adjusting analysis by these factors.  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985



986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026

		Centre			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%
Randomisation stratified by and adjustment made	Yes	17	21	81%	9	15	60%

**Table 5: Planned statistical adjustments through analysis in multi-centre and multi-surgeon trials**

1027  
1028  
1029 *3.9. Funder led considerations*  
1030

1031 *3.9.1. Commissioning briefs*  
1032  
1033

1034 Of the fifteen commissioning briefs, one permitted single centre studies and one required a multi-  
1035 centre setting. No other brief gave guidance with respect to number of centres. Two briefs identified  
1036 surgical learning considerations as an issue to address: the first indicating outcomes may be  
1037 independent of surgeon grade and the second:  
1038  
1039  
1040  
1041

1042 *“Proposals should account for the possibility of a learning curve affecting the outcomes of [surgery].”*  
1043  
1044

1045 *3.9.2. Changes driven by funder*  
1046  
1047

1048 Response to referee comments were available for 40 studies (n=40/49, 81.6%). Fourteen examples of  
1049 change within twelve applications were identified. Funder concerns led to sample size adjustment for  
1050 surgeon (n=3); randomisation balanced for surgeon (n=2) and centre (n=1); and improved  
1051 generalisability by increasing the number of centres (n=3):  
1052  
1053  
1054  
1055

1056 *“The Board suggested that the team should consider the addition of a second centre to demonstrate*  
1057 *generalisability and help with recruitment.”*  
1058  
1059

1060  
1061 In one application, funders requested applicants increase homogeneity in treatments and the  
1062 applicants argued against this.  
1063  
1064

1065 *“To ensure homogeneity in treatments we have consulted with our participating surgeons [and] the*  
1066 *National [...] Registry and agreed to specify the use of a CE marked [device...there are three main*  
1067 *devices]. Surgical trials that specify a single type of [device] are notoriously difficult to conduct and we*  
1068 *do not believe such a design could recruit surgeons, nor would the outputs be generalisable. “*  
1069  
1070  
1071  
1072  
1073  
1074

1075 Further considerations with regard to surgeon credentials (n=3) and the impact of surgeon equipoise  
1076 on recruitment (n=1) were also funder driven.  
1077  
1078

1079 *“The sample size has been increased from a total of [n] patients to a total of [1.4n] to take into account*  
1080 *clustering of surgeon as per the feedback from the first stage.”*  
1081  
1082  
1083  
1084  
1085

#### 4. CONCLUSIONS

This review has investigated the decision-making behind intended design and analysis of 50 randomised surgical trials funded by the NIHR EME and NIHR HTA programmes from 2012 to 2016. These results show frequent consideration of centres and surgeon impact during design, and these may be funder led, due to concerns around homogeneity or generalisability 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 are highlighted within reporting of non-pharmacological interventions guidelines, [22, 23] and a review of the published literature identified a deficiency in adherence to these [16]. In 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 multi-centre and 12% of multi-surgeon studies reported a statistical adjustment of these within published manuscripts. This was 43% and 68% respectively in this cohort. When randomisation was stratified by centre or surgeon, this was accounted for in the analysis in 30% of multi-centre and 40% of multi-surgeon trials in the published manuscripts, as oppose 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 recognised the need to increase research-based evidence in surgery. Applications were invited that evaluated technology-driven implanted or implantable medical devices, surgical procedures or surgical services. As a clinical trial is typically a major financial investment, [24] applicants need to assure funders that their proposal is important, well designed and demonstrates

1145  
1146  
1147 scientific value to add to the current evidence base. Each application undergoes a peer review process,  
1148  
1149 where 'experts' critically review the trial to ensure standards are met in terms of design, quality,  
1150  
1151 feasibility, acceptability and importance of the topic. [17, 18] A strength of this review is the insight  
1152  
1153 into the designs proposed to funders, and impact of feedback on subsequently funded studies.  
1154  
1155

1156  
1157 Whilst the degree of learning and clustering will vary trial-to-trial, many interventions require surgical  
1158  
1159 skill in their delivery regardless of whether or not the surgery is the intervention of interest. The impact  
1160  
1161 of any potential imbalance in delivery on comparing interventions should be considered at trial outset  
1162  
1163 routinely. Early and careful consideration will ensure that procedures are standardised as completely  
1164  
1165 as possible such that, in severe cases, the trial team can alleviate any doubts about homogeneity  
1166  
1167 raised by the medical community should the trial results be questioned. [12] These results indicate  
1168  
1169 funder awareness of this early consideration, with one of the two examples of balancing  
1170  
1171 randomisation surgeon following recommendation being in a trial where surgery was not the  
1172  
1173 intervention of interest.  
1174

1175  
1176 When interpreting these results, it is important to consider the limitations of this review. First, only  
1177  
1178 successful applications could be included due to confidentiality constraints. It is therefore not possible  
1179  
1180 to determine whether the management of learning and clustering contributes to the success of the  
1181  
1182 application. However, given that the application process consists of iterations whereby peer reviewers  
1183  
1184 are able to request that researchers address paucities in their application, it is unlikely that a promising  
1185  
1186 application, lacking in the appropriate considerations, would be deemed unsuitable for funding  
1187  
1188 outright. More likely, researchers would be given the opportunity to make these considerations during  
1189  
1190 this iteration process. Second, as part of this iterative review, it is possible that additional discussions  
1191  
1192 at the funder board meetings did not make it in to the comments fed back to applicants. This could  
1193  
1194 mean that funders raised these issues more frequently than this review suggests. Third, due to the  
1195  
1196 nature of the grant application process, the funder impact observed may be in part due to an increased  
1197  
1198 awareness of the reviewers involved. Fourth, this work has focussed on a single funding body that  
1199  
1200  
1201  
1202  
1203

1204  
1205  
1206 primarily supports UK based research. However, trials supported span a wide range of surgical  
1207  
1208 specialties and health care conditions and results from this review will be generalisable to other  
1209  
1210 funding bodies with a similar peer review process.  
1211

1212  
1213 Fundamental to trial design and analysis is understanding the objectives. While considerations relating  
1214  
1215 to clustering and learning effects are not widely reported in main trial publications, these results  
1216  
1217 indicate both funders and researchers consider these aspects in order to address a specific research  
1218  
1219 question. Such issues may have varying relevance depending on the overall design of the trial. A very  
1220  
1221 pragmatic study may deliberately include surgeons and centres of all types and have less emphasis on  
1222  
1223 expertise and learning, whereas the delivery of the intervention in more explanatory studies is critical  
1224  
1225 and requires consideration during design and analysis. Another approach to overcoming these issues  
1226  
1227 is to provide quality assurance of the intervention. Early work to develop methods to achieve this have  
1228  
1229 been developed and it is expected that this will expand in the future. [25] Furthermore, these results  
1230  
1231 provide insight into the promising role of the funder as a driver to improving the, long criticised,  
1232  
1233 surgical evidence base. The funder, who has influence over whether or not and how studies are carried  
1234  
1235 out and has been suggested as a driver for improving the quality of research during the period of  
1236  
1237 growth for surgical trials [3], can play a valuable role in ensuring that future trials do not have the  
1238  
1239 same shortfalls as those in the past.  
1240

## REFERENCES

- [1] Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*. 2002; 12(1), 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, Kolas A, Kehoe S, Lancaster G, Donovan J, Petrinic T, Agha R and Pennell C. Progress in clinical research and IDEAL. *Lancet*. 2018; 392: 88-94.
- [4] Ahmed AU, van der Sluis PC, Issa Y, et al. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. *Ann Surg*. 2013: 258 (2).
- [5] Masters J and Costa M. How to build a randomised controlled trial. And how to decide when it is appropriate. *Royal College of Surgeons Bulletin*. 2017; 99(6).
- [6] Blencowe N, Cook JA, Pinkney T, et al. Delivering successful randomized controlled trials in surgery: methods to optimize collaboration and study design. *Clinical Trials*. 2017; 14(2), 211-218.
- [7] Applied Health Research in Surgery (2012) [National Institute for Health Research web site]. Available at: <https://www.nihr.ac.uk/funding-and-support/documents/themed-calls/Surgery.pdf>. Accessed January 9th, 2019.
- [8] Meara JG, Leather JM, Hagander L, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *The Lancet*. 2015; 386: 569-624.
- [9] Hu Y, Edwards BL, Brooks KD, et al. Recent trends in National Institute of Health Funding for Surgery: 2003 to 2013. *Am J Surg*. 2015. 209(6): 1083-1089.
- [10] National Institute for Health Research Unit on Global Surgery (2017) [Global Surg web site]. Available at: <http://globalsurg.org/about/>. Accessed January 9th, 2019.
- [11] Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet*. 2009; 374 (9695).

- [12] ICH E9 Expert Working Group. Statistical Principles for Clinical Trials: ICH Harmonized Tripartite Guideline. *Stats in Medicine*. 1999; 18:1905-1942.
- [13] Cook JA, Ramsay CR and Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clinical Trials*. 2004; 1:421-427.
- [14] Roberts C and Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials*. 2005; 2: 152-162.
- [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, and Gamble C. Randomized trials involving surgery did not routinely report considerations of learning and clustering effects. *Journal of Clinical Epidemiology*. 2018. 107; 27-35.
- [17] Health Technology Assessment. [National Institute for Health Research web site]. Available at: <https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technology-assessment/>. Accessed January 9th, 2019.
- [18] Efficacy and Mechanism Evaluation. [National Institute for Health Research web site]. Available at: <https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/efficacy-and-mechanism-evaluation/>. Accessed January 9th, 2019.
- [19] National Institute for Health Research Journals Library. [National Institute for Health Research web site]. Available at: <https://www.journalslibrary.nihr.ac.uk/journals/>. Accessed March 8th, 2019.
- [20] 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 January 9th, 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.](#)
- [1922] ] Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*. 2008; 148(4):295-309.
- [239] Boutron I, Altman DG, Moher D, et al. CONSORT statement for randomized trials for nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Annals of Internal Medicine*. 2017; 167(1):40-47.
- [241] Clinical Trials Toolkit: Funding proposal [Clinical Trials Toolkit web site]. Available at: <http://www.ct-toolkit.ac.uk/routemap/funding-proposal/>. Accessed January 9th, 2019.
- [252] Blencowe NS, Mills N, Cook JA, Donovan JL, Rogers CA, Whiting P, and Blazeby JM. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. *Br J Surg*. 2016. 103(10):1377-84.
- [26] [Gspomer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, and Schmidli H. A practical guide to Bayesian group sequential designs. \*Pharm Stat. England\*; 2014;13\(1\):71-80.](#)
- [23] ~~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 January 9th, 2019.~~



## **Acknowledgements**

The authors would like to thank the National Institute for Health Research Evaluation, Trials and Studies Coordinating Centre (NETSCC) for collating and permitting access to the data used for this review.

## **Funding**

This research was funded by the National Institute for Health Research Doctoral Fellowship Programme (Ref: DRF-2015-08-082). EJC is funded through this Fellowship Programme. JMB and JAC are part 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); JMB is also supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. CG is part funded by the Medical Research Council North West Hub for Trials Methodology Research – MR/K025635/1.

## **Declaration of interest**

CG is a member of the National Institute of Health Research Efficacy and Mechanism Evaluation Funding Committee. All authors were in receipt of funding from the National Institute of Health during the cohort period. JAC, JMB, CG, and GB were named applicants on trials included in the cohort.

EJC holds an NIHR funded doctoral fellowship.

## **Department of Health disclaimer**

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.

**Author statement**

EJC participated in the study design, drafted the manuscript, established the data access agreement, developed the data extraction form, and extracted and analysed the data. CG participated in the study design, developed the data extraction form, analysed the data and drafted the manuscript. ARH extracted the data. GB, JMB and JAC participated in the study design, reviewed the data extraction form and contributed to manuscript development. All authors read and approved the final manuscript.

## APPENDIX A

### Contents of Appendix A

Supplement A1: Data extraction form

Figure A1: Flowchart of eligibility

Table A1: Trial design details

Table A2: Recruitment and randomisation

## **Supplement A1: Data extraction form**

The following details were extracted from eligible funding applications:

### **SECTION 1: Trial details**

- 1.1. Funding identifier (CATEGORICAL – EME / HTA)
- 1.2. Trial name (FREETEXT)
- 1.3. Number of randomized controlled trials in application (NUMERIC)
- 1.4. Lead institute region (CATEGORICAL – by COUNTY)
- 1.5. Funding start year (CATEGORICAL – 2012 / 2013 / 2014 / 2015 / 2016 / 2017)
- 1.6. Documents available for review
  - 1.6.1. Commissioning brief (BINARY – Yes / No)
  - 1.6.2. Project description (BINARY – Yes / No)
  - 1.6.3. Funder changes (BINARY – Yes / No)
  - 1.6.4. Protocol (BINARY – Yes / No)

### **SECTION 2: Design details**

- 2.1. Trial design (CATEGORICAL - *Cluster / Crossover / Parallel / Factorial / Stepped wedge / N-of-1 / Sequential*)
- 2.2. Number of trial arms (NUMERIC)
- 2.3. Use of pilot or feasibility in design
  - 2.3.1. Pilot study (BINARY – Yes / No)
  - 2.3.2. Feasibility study (BINARY – Yes / No)

### **SECTION 3: Intervention of interest**

- 3.1. Nature of surgery delivered (BINARY – *As an intervention / As part of patient pathway*)

3.2. If surgery delivered in as an intervention, what is the comparator (CATEGORICAL – *Surgery / Medical / Other*)

3.3. If surgery is delivered as intervention and is also a comparator, what is the nature of the surgical comparator? (CATEGORICAL – *Alternative surgical procedure / Change to a component of the same procedure / Same procedure delivered at different time points*)

3.4. If surgery is delivered as intervention and is also a comparator, was an expertise based design utilised? (CATEGORICAL – *Pure: professionals delivering only one intervention / Hybrid: some professionals could deliver both*)

#### **SECTION 4: Recruitment**

4.1. Number of countries (BINARY – *Multiple / Single*)

4.2. Number of centres (BINARY – *Multiple / Single*)

4.3. Number of surgeons (BINARY – *Multiple / Single*)

#### **SECTION 5: Randomisation**

5.1. Method of randomisation (CATEGORICAL – *Dynamic allocation / Block / Simple*)

5.1.1. If dynamic allocation, specify (BINARY – *Minimisation / Other*)

5.2. Allocation ratio (BINARY – *Equal / Unequal*)

5.3. Randomisation unit (BINARY – ~~*Minimisation-Individual*~~ / *Dyad / OtherCluster*)

5.4. Randomisation stratified (BINARY – *Yes / No*)

5.4.1. If randomisation stratified, stratified by country (BINARY – *Yes / No*)

5.4.2. If randomisation stratified, stratified by centre (BINARY – *Yes / No*)

5.4.3. If randomisation stratified, stratified by surgeon (BINARY – *Yes / No*)

#### **SECTION 6: Centre and surgeon credentials**

6.1. Credentials defined (BINARY – *Yes / No, not reported*)

6.2. Centre credentials (FREETEXT)

6.3. Surgeon credentials (FREETEXT)

## **SECTION 7: Outcomes**

7.1. Outcomes (FREETEXT)

## **SECTION 8: Statistical considerations**

8.1. Sample size considerations e.g. adjusting for ICC (FREETEXT)

8.2. Planned exploratory analysis e.g. differences in outcome between centres (FREETEXT)

8.3. Formal analysis e.g. adjusting models (FREETEXT)

## **SECTION 9: Funder led considerations**

9.1. Commissioning brief (FREETEXT)

9.2. Funder led changes (FREETEXT)

Figure A1: Flowchart of eligibility

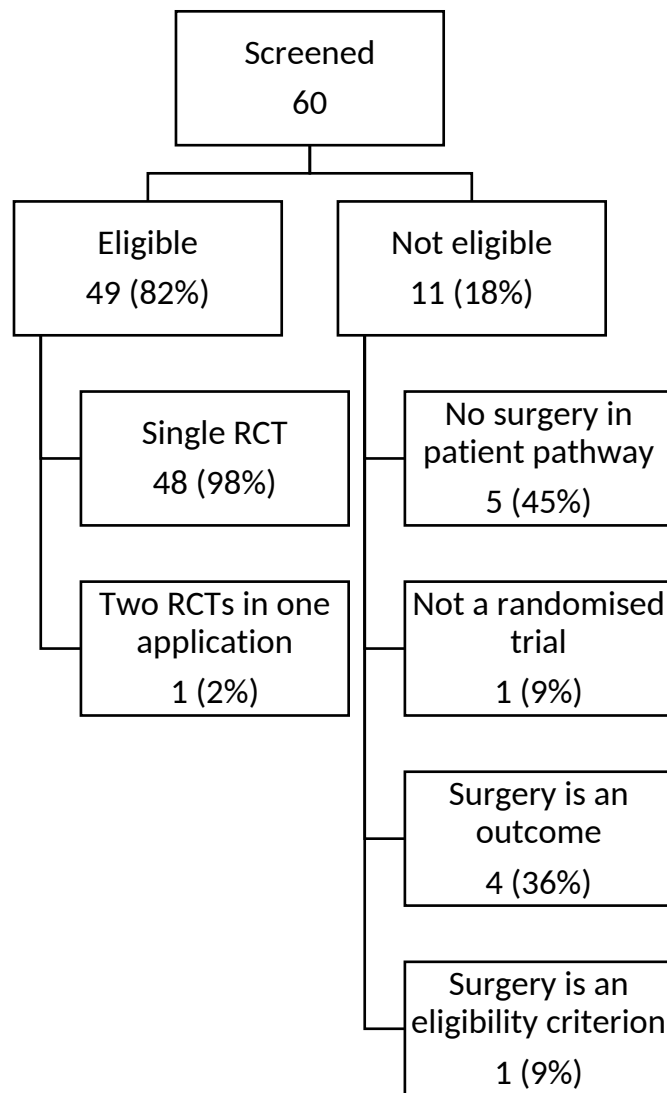


Table A1: Trial design details

Item	Category	n	N	n/N%
Type	Parallel	49	50	98%
	Sequential <a href="#">[25]</a>	1	50	2%
Number of trial arms	2	45	50	90%
	3	4	50	8%
	4	1	50	2%
Use of pilot or feasibility study, <a href="#">internal or external [230]</a>	Both pilot and feasibility	2	50	4%
	Pilot only	29	50	58%
	Feasibility only	11	50	22%
	No	8	50	16%
Nature of surgery delivered  If intervention comparator	As an intervention	39	50	78%
	As part of patient pathway	11	50	22%
	Surgery	21	39	54%
	Medical	7	39	18%
	Other	11	39	28%



Item	Category	n	N	n/N%
If surgical comparator	Alternative surgical procedure	13	21	62%
	Change to a component of the same procedure	6	21	29%
	Same procedure delivered at a different time point	2	21	10%

**Table A2: Recruitment and randomisation**

Item	Category	n	N	n/N%
Method of randomisation	Dynamic allocation	23	50	46%
	Minimisation	21	23	91%
	Other	2	23	9%
	Block	17	50	34%
	Not specified	10	50	20%
Allocation ratio	Equal	48	50	96%
	Unequal	1	50	2%
	Not specified	1	50	2%
Randomisation unit	Patient	50	50	100%
Randomisation stratified	Yes	46	50	92%
	No, not specified	4	50	8%
Multiple countries participating	Yes	3	50	6%
	No	45	50	90%
	Not reported	2	50	4%

Item	Category	n	N	n/N%
If yes, stratified by country	Yes	1	3	33%
	No	2	3	66%
Multiple centres participating  If yes, stratified by centre	Yes	49	50	98%
	No	1	50	2%
	Not reported	0	50	.
	Yes	28	49	57%
	No, justification provided	1	49	2%
	No, by other variables	17	49	35%
	No, not stratified	3	49	6%
Multiple surgeons participating  If yes, stratified by surgeon	Yes	22	50	44%
	No	0	50	.
	Not reported	28	50	56%
	Yes	8	22	36%
	No, justification provided	0	22	.
	No, by other variables	13	22	59%

Item	Category	n	N	n/N%
If yes, multicentre study	No, not stratified	1	22	5%
	Yes	21	22	96%
If yes, stratified by	No	1	22	5%
	Centre and surgeon	2	21	10%
	Centre, not surgeon	11	21	52%
	Surgeon, not centre	6	21	29%
	Neither centre nor surgeon	2	21	10%