Managing clustering effects and learning effects in the design and analysis of multicentre randomised trials:

A survey to establish current practice

Authors

Mrs Elizabeth J Conroy (EJC) MSc. (0000-0003-4858-727X) ejconroy@liv.ac.uk

Prof. Jane M Blazeby (JMB) MD. (0000-0002-3354-3330) j.m.blazeby@bristol.ac.uk

Dr. Girvan Burnside (GB) PhD. (0000-0001-7398-1346) g.burnside@liverpool.ac.uk

Assoc. Prof. Jonathan A Cook (JAC) PhD. (0000-0002-4156-6989)

Prof. Carrol Gamble (CG) PhD. (0000-0002-3021-1955) carrolp@liverpool.ac.uk

1 Department of Biostatistics, University of Liverpool, a member of Liverpool Health Partners, Liverpool, UK.

2 Centre for Surgical Research, Bristol Biomedical Research Centre, Population Health Sciences, University of Bristol, Bristol, UK.

3 Centre for Statistics in Medicine, University of Oxford, Oxford, UK.

4 North West Hub for Trials Methodology Research, University of Liverpool, Liverpool UK.

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

Name: Elizabeth J Conroy

Postal Address: Liverpool Clinical Trials 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
ABSTRACT

Background

Patient outcomes can depend on the treating centre, or health professional, delivering the intervention. Skills of health professionals improve with experience in delivery, meaning that outcomes may also be associated with changes in skill, or learning. Considering any potential difference in intervention delivery at trial design will ensure that any adjustments can be made, as appropriate, at the analysis. The objective of this work was to establish current practice for the allowance of clustering and learning effects in the design and analysis of randomised multicentre trials.

Methods

A ten-question survey was developed comprising open and closed questions that drew upon quotes from existing guidelines, references to relevant publications, and example trial scenarios. It was piloted and checked for face validity within the research team. All registered UK Clinical Research Collaborative registered Clinical Trials Units were invited to participate.

Results

Completed surveys were obtained from 44 of 50 Units. Clustering was managed through design by stratification, commonly by centre and less commonly by treatment provider. Most Units had allowed for learning by design through defining a minimum level of expertise for treatment provider (89%). One third of Units reported experience of expertise-based designs. The majority of Units indicated experience in adjusting for clustering during analysis, by centre or treatment provider, although approaches to doing so varied. Analysis of learning was rarely performed for the main analysis (n=1), although many Units reported approaches
to consider such effects, such as sensitivity analyses. Responders provided insight behind the approaches used within their Unit and reasons for, or against, alternative approaches.

**Conclusion**

This survey identifies widespread awareness of the potential methodological challenges associated with the design and analysis of multicentre trials, although approaches used and opinions on these vary. These results suggest that variation in approaches used exists both across Units and within, suggesting that this decision can depend on the type of trial being conducted. Reasons for approaches were provided and approaches justified by responders. These results highlight the need for more agreement between triallists about how to best design and analyse trials of different types and/or further research to establish optimal methods.

**Keywords:** trials, clinical trials unit, clinical trial, randomised controlled trial, complex intervention; surgical intervention; trial design; trial analysis; survey; clustering; learning
Background

Patient outcomes depend crucially on the treatment provider delivering the intervention. Where there is more than one treatment provider, outcomes observed in patients treated by the same treatment provider may be more similar than patients treated by other treatment providers, a phenomenon known as clustering. The potential for clustering is also present, albeit less obviously, for treating centre within a clinical trial. [1, 2] In addition to clustering, a change in skill in treatment delivery may be observed over time, specifically there may be a learning element experienced within one or all of the arms of the study observed during the course of the trial, meaning that trial outcomes may also change in be associated with changes in skill. [3] When comparing interventions within a clinical trial, it is imperative that any trial is designed under a common protocol, with regards to treatment delivery, and that the trial is conducted in accordance with this. At trial outset, a researcher may consider the homogeneity of any intervention under examination and the degree to which it is appropriate to standardise these procedures. [4] In extreme cases, where the trial results are questioned by the research community related to the study results, the trial team should be prepared to alleviate any doubts of heterogeneity of treatment effects. [5]

Difference in treatment delivery is often considered more of a concern in trials investigating a complex intervention, such as surgery. Trials involving a complex intervention are often criticised because of variability between intervention providers (clustering) but also due to
variability over time often as a result of increased experience (learning). Recognition, and management as appropriate, of clustering and learning is recommended, and it may have increased relevance within the surgical field, dependent upon the interventions being investigated and their routine use. Considering these aspects at trial outset will ensure that any necessary adjustment, to the design or analysis of the study, is applied in a manner appropriate for the intervention under investigation and support clinical decision making.

Whilst the notion of clustering and learning is familiar to many statisticians, the extent to which these considerations are made, and how, is unknown. A survey to establish current practice for the statistical management of clustering and learning effects in the design and analysis of randomised multicentre trials was undertaken within UK Clinical Research Collaborative registered Clinical Trials Units. This survey aimed to ascertain UK wide experience of running multicentre studies, in particular those investigating a complex or surgical intervention. In addition to establishing awareness of design issues associated with these studies and levels of concerns around these issues.

**Methods**

The survey was delivered at the bi-annual Statisticians Operational Group Meeting in April 2018. Attendees were statistical representatives from each of the UK Clinical Research Collaboration (UKCRC) registered Clinical Trials Units (CTUs). Units that did not have a representative present at this meeting, or did not respond, were contacted via email following the event and invited to participate. Registered Units were identified from the Network website on the 4th January 2018 (n=51, of which 50 were registered at time of survey, Supplementary Box 1).

**Ethics and Consent to Participate**
As the survey involved professionals and involved discussions of current practice, no formal ethical approval were deemed necessary. Consent was assumed upon participation, responders were free to refuse participation without providing a reason.

**Survey**

EJC and CG developed the survey and GB, JB and JAC reviewed and provided feedback. The survey was subsequently piloted and revised prior to roll out (*Supplementary Box 2*).

This survey was developed to establish experience in multicentre trials, in particular those investigating a complex intervention. Questions drew upon quotes from existing guidelines, references to relevant publications, and example scenarios developed by the study team ([4, 6, 10], *Box 1*). Questions included concepts such as Units experience in adjusting for clustering (therapist/surgeon or centre) or time varying effects (learning curves) and, when a Unit had experience, when and how adjustments are applied. This survey also aimed to establish awareness about design issues in surgery and levels of concern around these.

Questions were analysed and reported by Unit. To represent Unit practice and experience as a whole, Units with multiple responders were combined. However, due to the nature of the network meeting invites (one per registered CTU) multiple responders from a single CTU were minimised.

**Analysis**

Quantitative data from closed questions were analysed using descriptive statistics with standard statistical software [Statistical Analysis Software (SAS®) 9.1.4; SAS Institute Inc., Cary, NC, USA]; no formal statistical testing was undertaken.

EJC and CG identified themes within the free text answers, which were used to contextualise and illuminate quantitative responses.

To ensure anonymity, each Unit was assigned a project identification number.
Results

Unit participation and demographics

Forty-seven of the 50 UKCRC registered CTUs were represented at the network meeting on 28th April 2018. Of those present, 34 representatives from 31 Units (62%) participated. Following the meeting, Units without a completed survey were contacted, of which thirteen responded (n=13/19). Supplementary Table 1 provides further detail. The overall participation rate of registered Units was 88% (n=44/50). One representative from a newly registered Unit reported lack of experience as a reason for non-participation, reasons were not provided from the remaining five Units.

All responders had a statistical background with the majority of responders holding a senior or lead at their Unit (senior statistician: n=15/44, 34%; statistical lead: n=13/44, 30%). Supplementary Table 2 provides further detail.

Units listed on the UKCRC Resource Finder [11] as conducting cluster or surgical trials had participation rates 94% (n=16/17) and 92% (n=33/36) respectively (Supplementary Table 3).

Units with a methodological research area in complex interventions participated with a rate of 90% (n=35/39). The statistical roles of three-quarters of Units indicated experience in running trials with a complex intervention (n=32/44, 73%) and two-thirds in running trials with a surgical intervention (n=29/44, 66%), with twenty-five (57%) indicating experience in both. Seven Units stated that their Unit did not have experience in running trials with either type of intervention (n=7/44, 16%). One did not respond to this question (Question 1, Supplementary Table 4).

Managing effects through design

Clustering

Twenty-five Units had undertaken multicentre trials that did not stratify by centre (n=25/44, 57%, Question 2, Table 1 and Table 2). Common reasons for not stratifying by centre were
many centres with few participants (n=19/25, 76%) and expected homogeneity of treatment effect (n=11/25, 44%). Additional reasons for not stratifying by centre included allocation concealment in an open trial; logistical reasons; and grouping centres by region. One responder clearly indicated that this decision was influenced by the nature of the intervention stating:

“…drug trials less effect due to centre compared to say complex or surgical interventions.” [ID23]

One responder did stratify all their trials by centre alluded to concerns regarding potential for unequal distribution of costs across centres:

“This subject gets a lot of academic debate in some academic circles. But: our randomisation defaults to stratifying by centre; need to balance resources – don’t want to give one too many overheads; balancing avoids confounding; other opinions, such as Torgerson, exist.” [ID8]

Question 3 asked responders to consider five scenarios (Box 1, Table 1 and Supplementary Table 5), in particular their approach to stratifying the randomisation in trials of each type ran by their Unit. Responses to Scenario A, of which 39 Units had experience, indicated that most Units when running a trial with a large sample size, with multiple treatment providers per centre each recruiting a minimum of 10 participants, would stratify by centre alone (n=31/39, 87%).

Three would stratify by treatment provider alone (n=3/39, 8%). Seventy percent had experience of running trials like Scenario B, which was the same as Scenario A, only with a small sample size (n=31/44, 70%). As with Scenario A, most Units ran such trials by stratifying by centre alone (n=24/31, 77%) and few by treatment provider alone (n=2/31, 6%).

Responders had less experience running Scenario C trials, trials recruiting in several centres where treatment providers treated patients across centres (n=16/44, 36%). Again, most
common was stratification by centre only (n=14/16, 88%), with a greater number of Units indicating that they had stratified such trials by treatment provider only (n=3/16, 19%).

Units with experience running trials in Scenario D, trials recruiting from multiple centres, each with multiple treatment providers, that investigated a surgical intervention (n=25/44, 57%), also primarily stratified by centre only (n=21/25, 84%). One-fifth indicated stratifying by both centre and treatment provider in such trials (n=5/25, 20%).

Whilst Units had less experience running trials like Scenario E, which was similar to Scenario D but investigating substantially different interventions, stratification approaches were similar to Scenario D (Centre only: 13/16, 81%; both centre and treatment provider: 2/16, 13%).

Twelve responders provided free text explaining their approaches for stratification in each of the scenarios (Question 3, Supplementary Table 5). Two-thirds (n=8/12, 67%) commented on the feasibility of stratifying by treatment provider. Reasons were: concerns that there would be too few per strata [ID8, ID15, ID39]; treatment provider not known in advance [ID8, ID32]; delivered by a subset of treatment deliverers [ID1, ID39]; data not collected on treatment provider [ID13]; treatment differences assumed to be differences in facilities and protocols [ID17]; usually comparing the intervention policy and not the different aspects of the intervention [ID32]; treatment provider can change during the trial [ID30].

Other responses provided examples of stratification levels e.g. centre as hospital and treatment provider as operating surgeon [ID10]; two that this was trial specific [ID14, ID29]. One raised concerns with stratifying by centre:

“Recent conversions between senior statisticians advocate not stratifying by centre in any situation. They cited concerns regarding prediction of allocation.” [ID18]

When comparing stratification approaches across scenarios within Units (Question 3, Table 1), nineteen Units used the same approach across all scenarios they had experience in and
twenty changed their approach depending on the trial scenario (same: n=19/44, 43%;
different: n=20/44, 46%). Five had no experience in any of the suggested scenarios or did
not respond to the question.

Learning

The majority of responders (n=39/44, 89%) indicated they had accounted for learning by
defining a minimum level of expertise for treatment providers (Question 4, Table 1). Common
definitions were set in terms of delivering the trial intervention (n=31/44, 70%); treating the
condition within the patient population (n=24/44, 55%); and setting a minimum professional
level for treatment providers (n=22/44, 50%). Three delegated this responsibility to the
clinical investigators on the study. Examples of alternative approaches to specifying
minimum levels of expertise included: use of a surgical manual with senior surgeons signing
off treatment deliverers [ID16] and treatment deliverers being required to pass both surgical
and radiotherapy quality assurance [ID18].

Thirty percent of Units had used an expertise based trial design, in which participating
treatment providers provide only the intervention in which they have expertise (n=13/44,
30%, Question 5, Table 1).

Managing effects through analysis

Clustering

In trials stratified by centre, 55% of Units had subsequently adjusted by this stratification
factor in the analysis (n=24/44, 55%, Question 6, Table 3 and Table 4). This had been done
either by pre-specified grouping rules at the design stage (n=19/24, 83%); by an ad-hoc
approach (n=14/24, 58%); or by other approaches: grouped centres where numbers are
small [ID7, ID15]; site as a fixed effect [ID8]; or:

“Depends. Either include as a stratifying factor (small number of centres, large patient
numbers) or by including centre or treatment provider as a cluster.” [ID32]
Regardless of stratification approach used, very few Units had never adjusted for centre in the statistical model when comparing treatment \((n=3/44, 7\%, \text{ Question 7, Table 3 and Supplementary Box 3})\). Responders from Units that did \((39/44, 89\%)\), did so using fixed effects \((n=11)\); random effects \((n=12)\); or, depending on the circumstance, used either \((n=14)\). Two did not respond. Reasons in favour for fixed effects were ease of interpretation and less assumptions associated with it, [ID27]; and random effects as:

“Usually an underlying assumption that centre may be a surrogate for socioeconomic factors that may affect outcome and/or treatment effect and so often not happy to assume that there is an equal fixed treatment effect across all sites.” [ID16]

In trials stratified by treatment provider, 37% also subsequently adjusted the analysis \((n=16/44, 37\%, \text{ Question 6, Table 3 and Table 4})\). Three-quarters did so in accordance with pre-specified grouping rules \((n=12/16, 75\%)\) or using a more ad hoc approach \((n=7/16, 44\%)\).

Regardless of stratification approach used, 59% adjust for treatment provider in the statistical model when comparing treatment \((n=26/44, 59\%, \text{ Question 8, Table 3 and Supplementary Box 4})\). The majority of responders used a random effect \((n=18/26, 69\%)\), with one providing reason:

“If treatment provider was included as stratification factor it will be because we are concerned that the provider will have an impact on outcome but also because we would expect different population for different treatment providers.” [ID16]

When responders were asked to revisit the scenarios in Box 1, this time to consider investigating treatment by centre or treatment provider (Question 9, Table 3), exploring treatment by centre was universally most common across all scenarios. Exploring treatment by provider was rare. Twelve responders provided free text to explain their approaches for adjustment (Question 9, Supplementary Table 6). General themes for additional information provided were: that the decision is trial dependent [ID6, ID14]; concerns around sample size [ID6, ID7, ID39]; and, when explored, that this was informal. [ID5, ID8, ID14, ID32, ID38]
When comparing treatment interaction approaches across scenarios within Units (Question 9, Table 3), 24 Units used the same approach across all scenarios and twelve utilised a scenario specific approach (same: n=24/44, 56%; different: n=12/44, 27%). Eight had no experience in any of the suggested scenarios or did not respond to the question.

Seventy-three percent of Units explore heterogeneity by centre when a positive treatment effect is found (n=32/44, 73%, Question 10a, Table 3), whereas fewer explored heterogeneity by treatment provider (n=12/44, 27%, Question 10b, Table 3). Of those that do explore heterogeneity for either effect, the majority did so by graphical display (centre: n=31/32; treatment provider: n=11/12). Many also explored by analytical methods, for example significance testing (centre: n=22/32; treatment provider: n=9/12). Supplementary Tables 6 and 7 provides further detail.

Learning

Fifty-nine percent of Units included the treatment provider in the statistical model when comparing treatment (n=26/44, 59%), two of which had treated this as a time-varying covariate (Question 8, Table 3), with one specifying:

“Fairly crude by letting the number of procedures in the trial increase the relevant surgeon’s experience (ignoring procedures done outside of the trial of course!)” [ID38]

Those that had not used a time varying effect had experience of exploring learning through a sensitivity analysis [ID35] or secondary analyses [ID8, ID39] to check for learning effect exploring learning effects with neither being significant. The latter adding that:

“Had we found evidence of learning, we would have had awkward additional data summaries and presentations”

Two responders had not considered such analyses [ID7, ID23] and one provided time restrictions as a reason for not doing so [ID30].

Discussion
This survey identifies that despite multicentre trials being prominent across all Units, there is a UK-wide variation of designing and analysing these trials with respect to clustering and learning effects. Approximately half of Units changed their approach to design and analysis when presented with five example trial scenarios, each with varying levels of complexity such as small sample size per centre and complex interventions, such as surgery. This finding suggests that variation can exist both across Units and within, suggesting that this decision can depend on the type of trial being conducted. Units indicate awareness of the potential methodological challenges associated with the design and analysis of multicentre trials, although approaches used and opinions on these vary. The high response rate achieved provides insight into the general and current practice of managing clustering and learning effects in multicentre trials investigating varying types of intervention. They indicate the need for a more unified approach to the design and analysis of trials where outcomes are associated with the delivery of the intervention and/or more research in this field.

When adjusting for clustering within the design, a higher proportion than expected ran trials that did not stratify by centre (52%). Most commonly, this was due to too many centres and not enough participants within centre. Stratifying by centre was most common in all scenarios, while stratifying by treatment provider was consistently rare but more common in trials with a surgical intervention. Stratifying by treatment provider raised pragmatic concerns e.g. provider not known pre randomisation, or concerns over relevance to research question. Half of responders had adjusted by centre following stratifying by the same, most commonly this was done by pre-specified grouping rules established at the design stage or using an ad hoc approach determined after design due to small numbers per group. Regardless of stratification approach, eight tenths of responders had adjusted for centre in the statistical model. There were mixed opinions on how this adjustment was made i.e. by fixed or random effects with reasons provided for and against both approaches. When a positive treatment effect is found, three quarters and one third stated that they then explore
heterogeneity by centre and treatment provider respectively, all did so using graphical displays.

Adjusting for learning by design through defining a minimum level of expertise for health professionals participating in the trial [4] was most common, with almost all responders (89%) applying these to studies within their Unit. Less than a third indicated experience in conducting expertise based designs, a design that can be particularly useful when comparing substantially different interventions. This finding suggests these designs are more commonly implemented than suggested by the literature [8, 12] Common concerns around adjusting and exploring experience were that the results may be “awkward” and that “care should be taken” in their presentation.

Strengths of this study were that although this survey was limited to registered Units, responders represent wide geographic coverage within the United Kingdom, spanning a diverse range of medical conditions and associated methodologies. In addition, participating Units are known to comply with required regulatory standards and meet acceptable standards of quality. [13] All responders were experienced trialists who either were Statistical Lead at their Unit or a nominated Statistical representative. Publicly funded trials cover a diversity of interventions [14] and are generally not seeking a marketing authorisation from the competent authorities and this may impact the approaches taken in line with heterogeneity of effects by cluster or time. Limitations of this work are that it represents statistical practice within the UK in leading trial centres, with global practice unknown. However, the survey drew upon internationally accepted guidelines [4] for best practice and therefore the opinions and experiences are applicable beyond the UK. Second, some of the observed responses may have related to the different types of surgical trials that the CTUs conduct. Not all surgical trials include interventions where there is learning, indeed, one would anticipate that most pragmatic large scale trials do not have ‘learning’ effects because they include interventions that are stabilised and in widespread use. Whilst the survey allowed for free-text responses, a more focussed survey, achieved using qualitative research
methods, would be needed to examine these issues. Third, the volume of studies designed by each Unit will vary widely, meaning that the experiences reported from larger units may not be indicative of all studies ever ran at that Unit. However, responders were able to complete the survey with additional support within their Unit if deemed appropriate.

Conclusions

This survey is the first to report on the experience and management approaches with regards to clustering effects and the learning curve in multicentre randomised trials. Importantly, responders, who were highly experienced in the design and analysis of such studies, appear to have awareness of when to make such considerations. Whilst approaches to management are varied, and this variation may be trial dependent within Unit, reasons for approaches reported were provided and approaches justified. Historically, guidance on the design, analysis and reporting of randomised controlled trials was developed more generally to support consistency in approaches across a more conventional randomised controlled trial [4, 14-15], with the development of more intervention specific guidelines being established following these to address the additional complexities across different types of trials. [6, 16-18] Intervention specific guidelines may have led to the variation and justifications identified in this survey. These results highlight the need for more agreement between triallists about how to best design and analyse trials of different types and/or further research to establish optimal methods.

LIST OF ABBREVIATIONS

UKCRC  United Kingdom Clinical Research Collaboration

CTU    Clinical Trials Unit
DECLARATIONS

Ethics approval and consent to participate
As the survey involved professionals and involved discussions of current practice, no formal ethical approval were deemed necessary. Consent was assumed upon participation, responders were free to refuse participation without providing a reason.

Consent for publication
Not applicable.

Availability of data and materials
The anonymised datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare there is no conflict of interest.

Funding
This research was funded by the National Institute for Health Research (NIHR) Doctoral Fellowship Programme (DRF-2015-08-082). EJC is funded through this Fellowship Programme. JMB is supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol (BRC-1215-20011). CG is part funded by the Medical Research Council North West Hub for Trials Methodology Research (MR/K025635/1) and the University of Liverpool. JMB is an NIHR Senior Investigator.

**Authors’ contributions**

EJC participated in the study design, drafted the manuscript, developed the survey, distributed the survey in person, and extracted and analysed the data. CG participated in the study design, developed the survey, distributed the survey by email, analysed the data and drafted the manuscript. GB, JMB and JAC participated in the study design, developed the survey and contributed to manuscript development. All authors read and approved the final manuscript.

**Acknowledgements**

The authors would like to thank all representatives from participating UK Clinical Research Collaborative registered Clinical Trials Units for completing the survey.

**Department of health disclaimer**

The views expressed are those of the authors and not necessarily of the National Health Service, the National Institute for Health Research (NIHR) or the Department of Health and Social Care.
REFERENCES

[1] Lee KJ, and Thompson SG. Clustering by health professional in individually randomised trials. BMJ. 2005;330:142.

[2] Cook JA, Bruckner T, MacLennan GS, and Seller CM. Clustering in surgical trials - database of intracluster correlations. Trials201213:2

[3] Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning effects in surgical trials. Clinical Trials. 2004;1(5):421-427

[4] ICH E9 Expert Working Group. Statistical Principles for Clinical Trials: ICH Harmonized Tripartite Guideline. Stats in Medicine. 1999; 18:1905-1942.

[5] Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, and Gamble C. Funders improved the management of learning and clustering effects through design and analysis of randomised trials involving surgery. Journal of Clinical Epidemiology. 2019; 446:28-35.
[6] Developing and evaluating complex interventions. [Medical Research Council website.] Available at: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/. Accessed January 23rd, 2019.

[7] Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet.* 2009; 374 (9695).

[8] 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.* 2019; 107:27-35.

[9] Roberts C and Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials.* 2005; 2: 152-162.

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

[11] UKCRC Registration ID Numbers. [United Kingdom Clinical Research Collaboration website.] Available at: https://cdn.ymaws.com/www.ukcrc-ctu.org.uk/resource/resmgr/registration_ids/2018-19_reg_ids_oct18.pdf. Accessed May 9th, 2018.

[12] Cook JA, Campbell MK, Gillis K, and Skea Z. Surgeons’ and methodologists’ perceptions of utilising an expertise based randomised controlled trial design: a qualitative study. *Trials.* 2018. 19: 478.

[13] McFadden E, Bashir S, Canham S, et al. The impact of registration of clinical trials units: the UK experience. *Clinical Trials.* 2015; 12: 166-173.

[14] ICH Harmonised Tripartite Guideline: Guideline for Good Statistical Practice E6(R1). Available at: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf. Accessed May 13th, 2019.
ICH Harmonised Tripartite Guideline: Structure and content of clinical study reports E3. Available at: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf. Accessed May 13th, 2019.

Boutron I, Moher SD, Altman DG, Schulz KF, Ravaud P. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacological treatments. Ann Intern Med. 2008;148:60–6.

Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P; CONSORT NPT Group. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. Ann Intern Med. 2017;167(1):40-47.

Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ. 2008; 337:a2390.

**TABLE AND FIGURES**

**Box 1: Example trial scenarios**

| Scenarios |
|-----------|
| A | A trial with a large\(^1\) sample size, recruiting in several centres each with multiple treatment providers. |
| B | A trial with a small\(^2\) sample size, recruiting in several centres, each with multiple treatment providers. |
| C | A trial that recruit within several centres, where treatment providers treat patients across recruiting centres i.e. treatment provider is not unique to a centre. |
| D | A trial recruiting from several centres, each with multiple treatment providers, investigating a surgical intervention. |
| E | A trial recruiting from several centres, each with multiple treatment providers, investigating substantially different surgical interventions e.g. a trial comparing surgery to an injection. |

\(^1\) Centres recruiting at least ten patients per site; \(^2\) Centres recruiting 2 to 3 patients
Table 1: Methods to managing clustering and learning by design

| Question                                                                 | Category       | Response statistics |
|-------------------------------------------------------------------------|----------------|---------------------|
| Does your Unit have any multicentre trials that do not stratify randomisation by centre? | Yes, No, No response | n = 25, N = 44, n/N = 57% |
|                                                                         |                |                     |
| In each of the following scenarios, how was the randomisation stratified in trials that your Unit has run? | Experience in trial type, Centre, Treatment provider, Both, Neither | n = 39, N = 44, n/N = 89% |
|                                                                         |                |                     |

(1) Recruiting in several centres, each with multiple treatment providers.
| Question                                                                 | Category                                                                 | n  | N  | n/N % |
|-------------------------------------------------------------------------|--------------------------------------------------------------------------|----|----|-------|
| No response                                                             |                                                                          | 1  | 44 | 2%    |
| B  Small sample size, \(^{(2)}\) recruiting in several centres, each with multiple treatment providers | Experience in trial type                                                 |    |    |       |
|                                                                          | Centre                                                                   | 24 | 31 | 77%   |
|                                                                          | Treatment provider                                                       | 2  | 31 | 6%    |
|                                                                          | Both                                                                     | 2  | 31 | 6%    |
|                                                                          | Neither                                                                  | 7  | 31 | 23%   |
|                                                                           | No experience in trial type                                              | 12 | 44 | 27%   |
|                                                                           | No response                                                              | 1  | 44 | 2%    |
| C  Recruiting in several centres, where treatment providers treat patients across recruiting centres (treatment provider is not unique to a centre) | Experience in trial type                                                 |    |    |       |
|                                                                          | Centre                                                                   | 14 | 16 | 88%   |
|                                                                          | Treatment provider                                                       | 3  | 16 | 19%   |
|                                                                          | Both                                                                     | 1  | 16 | 6%    |
|                                                                          | Neither                                                                  | 0  | 16 | 0%    |
|                                                                           | No experience in trial type                                              | 27 | 44 | 61%   |
| Question                                                                 | Category                                                                 | n  | N  | n/N % |
|-------------------------------------------------------------------------|--------------------------------------------------------------------------|----|----|-------|
|                                                                         | No response                                                              | 1  | 44 | 2%    |
| d A trial investigating a surgical intervention, recruiting from several centres, each with multiple treatment providers | Experience in trial type                                                 | 25 | 44 | 57%   |
|                                                                         | Centre                                                                   | 21 | 25 | 84%   |
|                                                                         | Treatment provider                                                       | 3  | 25 | 12%   |
|                                                                         | Both                                                                     | 5  | 25 | 20%   |
|                                                                         | Neither                                                                  | 3  | 25 | 12%   |
|                                                                         | No experience in trial type                                              | 17 | 44 | 39%   |
|                                                                         | No response                                                              | 2  | 44 | 5%    |
| e Recruiting from several centres, each with multiple treatment providers, comparing substantially different interventions e.g. surgery to an injection | Experience in trial type                                                 | 16 | 44 | 36%   |
|                                                                         | Centre                                                                   | 13 | 16 | 81%   |
|                                                                         | Treatment provider                                                       | 0  | 16 | 0%    |
|                                                                         | Both                                                                     | 2  | 16 | 13%   |
|                                                                         | Neither                                                                  | 2  | 16 | 13%   |
|                                                                         | No experience in trial type                                              | 26 | 44 | 59%   |
| Question | Category | n | N | n/N% |
|----------|----------|---|---|------|
| 3        | No response       | 2 | 44 | 5%   |
|          | In scenarios where Unit has experience, approaches to stratification changes across scenario i.e. within Unit variation to stratification | Different approaches across scenarios | 20 | 44 | 46% |
|          |                      | Same approach across all scenarios | 19 | 44 | 43% |
|          |                      | No response to Question 3 | 5 | 44 | 11% |
| 4        | In the trials ran by your Unit, have you defined a minimum level of expertise for the health professionals participating in the trial in terms of: | Treating the condition within the patient population | 24 | 44 | 55% |
|          |                      | Delivering the trial intervention | 31 | 44 | 70% |
|          |                      | Setting a minimum professional level of treatment providers | 22 | 44 | 50% |
|          | Other approach:     | Based on paramedic experience (defined by years in service) | 1 | 44 | 2% |
|          |                      | Based on surgeon experience (at or beyond a certain level) | 1 | 44 | 2% |
| Question                                                                 | Category                                                                 | n | N  | n/N % |
|------------------------------------------------------------------------|---------------------------------------------------------------------------|---|-----|-------|
| Centre required to conduct a certain number of operations per year.    |                                                                                         | 1 | 44 | 2%    |
| Clinical decision for Chief Investigator                                |                                                                                         | 1 | 44 | 2%    |
| Deliverer required to pass surgical and radiotherapy quality assurance  |                                                                                         | 1 | 44 | 2%    |
| Depends on phase of trial – early or pragmatic require different levels  |                                                                                         | 1 | 44 | 2%    |
| In our stepwise study, all therapists were experienced but the intervention was brand new. |                                                                                         | 1 | 44 | 2%    |
| Investigators who define research question are experts in the field and have trained staff to deliver intervention |                                                                                         | 1 | 44 | 2%    |
| No consistent approach across all our studies.                          |                                                                                         | 1 | 44 | 2%    |
| No unit wide policy – decided trial by trial depending on intervention and setting |                                                                                         | 1 | 44 | 2%    |
| Question                                                                 | Category                                                                                                                                                                                                 | n  | N  | n/N% |
|-------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|----|------|
| Has your Unit conducted trials with an expertise-based design, in which  | Surgeon manuals signed off by ‘senior’ surgeon prior to participation                                                                                                                                     | 1  | 44 | 2%   |
| participating treatment providers provide only the intervention to which | Surgical team led by consultant, who submits video measured for quality assurance, prior to participation.                                                                                               | 1  | 44 | 2%   |
| they have expertise?                                                    | These have been implicitly taken as a Chief Investigator and Principal Investigator                                                                                                                       | 1  | 44 | 2%   |
|                                                                         | Training provided to health care professionals in order to participate                                                                                                                                   | 1  | 44 | 2%   |
|                                                                         | No, or no response                                                                                                                                                                                        | 5  | 44 | 11%  |
| Has your Unit conducted trials with an expertise-based design, in which  | Yes, when applicable\(^3\)                                                                                                                                                                                | 13 | 44 | 30%  |
| participating treatment providers provide only the intervention to which | No, with justification                                                                                                                                                                                   | 1  | 44 | 2%   |
| they have expertise?                                                    | No                                                                                                                                                                                                       | 26 | 44 | 59%  |
|                                                                         | No response                                                                                                                                                                                              | 4  | 44 | 9%   |
Notes: (1) With centres each recruiting at least ten patients; (2) With centres each recruiting 2-3 patients; (3) We only have one grant application which we’ve proposed an expertise bases design this year but no prior experience of running a trial with such a design before. [ID22]
Table 2: Reasons for having multicentre studies that do/do not stratify by centre (Question 2)

| Unit has multicentre trials that do not stratify randomisation by centre? | Yes (N=25) | No (N=18) |
|---|---|---|
| **Reason(s) provided** | | |
| Expected homogeneity of treatment effect across centres | 11 (44%) | 2 (11%) |
| No interest in centre effect | 4 (16%) | 1 (6%) |
| Lots of centres with few participants per centre | 19 (76%) | 1 (6%) |
| Not convinced of appropriateness of either fixed or random effect models for centres in the trial | 1 (4%) | 0 (0%) |
| **Other reason provided** | | |
| Aids in blinding if trial open label | 1 (4%) | 0 (0%) |
| Balance against other important factors. Centre effect less important in drug trials compared to complex or surgical interventions | 1 (4%) | 0 (0%) |
| Concern that in an unblinded trial, stratifying by centre would make it easier to predict the treatment allocated to the next patient (Kahan et al. Trials (2015) 16:405). | 1 (4%) | 0 (0%) |
| For practical reasons | 0 (0%) | 1 (6%) |
| Intervention takes place out of hospital. | 1 (4%) | 0 (0%) |
Large sample size with small/moderate number of centres. We expect balance to be achieved with simple randomisation.

|   |   |   |   |
|---|---|---|---|
|   | 1 | 4% | 0 | 0% |
| Likely to stratify by geographical region if not by centre. | 1 | 4% | 0 | 0% |
| Randomisation system defaults to stratifying by centre but one example where minimised trial did not. Need to consider balance of resources and avoid confounding. There is a lot of academic debate. See Torgerson. | 0 | 0% | 1 | 6% |
| Sometimes stratify by region | 1 | 4% | 0 | 0% |
| Stratified by treatment provider within centres and treatment providers unique within centre. | 1 | 4% | 0 | 0% |
| Undertaken in limited/exceptional circumstances only e.g. feasibility studies. | 1 | 4% | 0 | 0% |
Table 3: Methods to managing clustering and learning by analysis

| Question | Category | n | N | n/N | % |
|----------|----------|---|---|-----|---|
| 6 a      | Assuming that you have stratified by centre, do you combine by the stratification factor for the purpose of analysis? If so how. | Yes | 24 | 44 | 55% |
|          |          | Pre-specified grouping rules at design stage | 19 | 24 | 83% |
|          |          | Ad hoc approach e.g. determined after design due to small numbers per group | 14 | 24 | 58% |
|          |          | Other grouping rule or further details provided | 6 | 24 | 26% |
|          | No       | 17 | 44 | 39% |
|          | No response | 3 | 44 | 7% |
| 6 b      | Assuming that you have stratified by treatment provider, do you combine by the stratification factor for the purpose of analysis? If so how? | Yes | 16 | 44 | 37% |
|          |          | Pre-specified grouping rules at design stage | 12 | 16 | 75% |
|          |          | Ad hoc approach e.g. determined after design due to small numbers per group | 7 | 16 | 44% |
|          |          | Other grouping rule or further details provided | 5 | 16 | 31% |

See Table 4 for further details.
| Question | Category                                      | n  | N  | n/N | %  |
|----------|----------------------------------------------|----|----|-----|----|
| 7        | Does your Unit include centre in the         | 39 | 44 | 89% |    |
|          | statistical model when comparing treatment?  | 18 | 39 | 46% |    |
| a        | Yes, and assuming that the sample size       | 6  | 39 | 15% |    |
|          | allows either, would you treat this          | 15 | 39 | 38% |    |
|          | effect as fixed or random?                   | 3  | 44 | 7%  |    |
|          | See Supplementary Box 3 for                  | 2  | 44 | 5%  |    |
|          | But only if it was used to stratify          | 14 | 39 | 36% |    |
|          | randomisation                                |    |    |     |    |
|          | Always                                       | 11 | 39 | 28% |    |
|          | Sometimes<sup>(1)</sup>                      | 12 | 39 | 31% |    |
|          | No, never                                    |    |    |     |    |
|          | No response<sup>(2)</sup>                    |    |    |     |    |
|          | Fixed or random, depending on circumstances  |    |    |     |    |
|          | Fixed                                        |    |    |     |    |
|          | Random                                       |    |    |     |    |
|          | No response                                  |    |    |     |    |
| Question | Category | n | N | n/N % |
|----------|----------|---|---|-------|
| 8        | Does your Unit include treatment provider in the statistical model when comparing treatment? | Yes | 26 | 44 | 59% |
|          | But only if it was used to stratify randomisation | 8 | 26 | 31% |
|          | Always | 0 | 26 | 0% |
|          | Sometimes$^{(3)}$ | 18 | 26 | 69% |
|          | No, never | 13 | 46 | 30% |
|          | No response$^{(4)}$ | 5 | 44 | 11% |
| a        | If yes, and assuming that the sample size allows either, would you treat this effect as fixed or random? | Fixed or random, depending on circumstances | 4 | 26 | 15% |
|          | Fixed | 2 | 26 | 8% |
|          | Random | 18 | 26 | 69% |
|          | No response | 2 | 26 | 8% |
| Question | Category | n | N | n/N % |
|----------|----------|---|---|-------|
| B If yes, has this effect ever been treated as time varying within the statistical model? | Yes | 2 | 26 | 8% |
| | No | 21 | 26 | 81% |
| | No response | 3 | 26 | 12% |
| 9 In each of the following scenarios, regardless of the randomisation stratification approach, has a treatment by centre or surgeon interaction investigated, in trials that your Unit has run? Select all that apply. | Experience in trial type | Centre | 16 | 35 | 46% |
| | Treatment provider | 4 | 35 | 11% |
| | Both | 3 | 35 | 9% |
| | Neither | 20 | 35 | 57% |
| | No experience in trial type | 7 | 44 | 16% |
| | No response | 2 | 44 | 5% |
| Question                                                                 | Category                                    | n  | N  | n/N % |
|-------------------------------------------------------------------------|---------------------------------------------|----|----|-------|
| b Small sample size, \(^6\) recruiting in several centres, each with multiple treatment providers | Experience in trial type                   | 30 | 44 | 68%   |
|                                                                         | Centre                                      | 5  | 30 | 17%   |
|                                                                         | Treatment provider                          | 0  | 30 | 0%    |
|                                                                         | Both                                        | 0  | 30 | 0%    |
|                                                                         | Neither                                     | 25 | 30 | 83%   |
|                                                                         | No experience in trial type                 | 12 | 44 | 27%   |
|                                                                         | No response                                 | 2  | 44 | 5%    |
| c Recruiting in several centres, where treatment providers treat patients across recruiting centres (treatment provider is not unique to a centre) | Experience in trial type                   | 15 | 44 | 34%   |
|                                                                         | Centre                                      | 4  | 15 | 27%   |
|                                                                         | Treatment provider                          | 1  | 15 | 7%    |
|                                                                         | Both                                        | 0  | 15 | 0%    |
|                                                                         | Neither                                     | 11 | 15 | 73%   |
|                                                                         | No experience in trial type                 | 27 | 44 | 61%   |
|                                                                         | No response                                 | 2  | 44 | 5%    |
| Question                                                                 | Category                                              | n  | N  | n/N % |
|-------------------------------------------------------------------------|-------------------------------------------------------|----|----|-------|
| d  A trial investigating a surgical intervention, recruiting from several centres, each with multiple treatment providers | Experience in trial type                              | 21 | 44 | 48%   |
|                                           | Centre                                                | 5  | 19 | 24%   |
|                                           | Treatment provider                                    | 3  | 19 | 14%   |
|                                           | Both                                                  | 1  | 19 | 5%    |
|                                           | Neither                                               | 14 | 19 | 67%   |
|                                           | No experience in trial type                           | 19 | 44 | 43%   |
|                                           | No response                                           | 4  | 44 | 9%    |
| e  Recruiting from several centres, each with multiple treatment providers, comparing substantially different interventions e.g. surgery to an injection | Experience in trial type                              | 14 | 44 | 32%   |
|                                           | Centre                                                | 5  | 14 | 36%   |
|                                           | Treatment provider                                    | 1  | 14 | 7%    |
|                                           | Both                                                  | 0  | 14 | 0%    |
|                                           | Neither                                               | 9  | 14 | 64%   |
|                                           | No experience in trial type                           | 26 | 44 | 59%   |
## Question 3: Exploration of Heterogeneity

| Question                                                                 | Category                                                                 | Response statistics |
|--------------------------------------------------------------------------|--------------------------------------------------------------------------|---------------------|
| In scenarios where Unit has experience, approaches to stratification changes across scenario i.e. within Unit variation to stratification | No response                                                               | n  | N  | n/N % |
| Different approaches across scenarios                                     | Same approach across all scenarios                                        | 12 | 44 | 27%   |
| No response to Question 3                                                 |                                                                         | 8  | 44 | 18%   |

10. a. If a positive treatment effect is found, does your Unit explore heterogeneity of treatment effects by centre?  

See Supplementary Table 7 for further details.

i. If yes to a, do you explore by graphical display?  

| Yes                                                                 | 32 | 44 | 73%   |
| No                                                                  | 9  | 44 | 20%   |
| No response                                                          | 3  | 44 | 7%    |

If yes to a, do you explore by graphical display?

| Yes                                                                 | 31 | 32 | 97%   |
| No                                                                  | 0  | 32 | 3%    |
| No response                                                          | 1  | 32 | 3%    |
| Question | Category | n | N | n/N % |
|----------|----------|---|---|-------|
| ii. If yes to a, do you explore by analytical methods e.g. significance testing? | Yes | 22 | 32 | 69% |
| | No | 5 | 32 | 16% |
| | No response | 5 | 32 | 16% |
| b) If a positive treatment effect is found, does your Unit explore heterogeneity of treatment effects by treatment provider? | Yes | 12 | 44 | 27% |
| | No | 23 | 44 | 52% |
| | No response | 9 | 44 | 20% |
| 1. If yes to b, do you explore by graphical display? | Yes | 11 | 12 | 92% |
| | No | 0 | 12 | 0% |
| | No response | 1 | 12 | 8% |
| ii. If yes to b, would you explore by analytical methods e.g. significance testing? | Yes | 9 | 12 | 75% |
| | No | 1 | 12 | 8% |

See Supplementary Table 8 for further details.
Question | Category | n | N | n/N %
--- | --- | --- | --- | ---
Testing? | No response | 2 | 12 | 17%

Notes: (1) “Sometimes” here is “usually” – it is a rare exception where we don’t. [ID10]; (2) No Standard Operating Procedure in place. [ID3]; (3) “Sometimes” here is “usually” – it is a rare exception where we don’t. [ID10]; (4) No experience in trials of this type. [ID1] Not applicable. [ID2]; (5) With centres each recruiting at least ten patients; (6) We only have one grant application which we’ve proposed an expertise bases design this year but no prior experience of running a trial with such a design before. [ID22].

Table 4: Other grouping rules when randomisation is stratified by (a) centres or (b) treatment providers

(Question 6)

Centre stratified:

ID4  Would normally analyse together but adjust for stratification factors (which normally include centres) in analysis.

ID7  There will be instances where we have combined centres at the analysis stage due to small numbers.
Different statisticians/trials do different things. Often site=fixed effect and course within site = random effect. If too few within site then would combine.

Retain structure at analysis.

Have grouped by region / country where numbers are small. Any adjustment should be documented in SAP and final decision regarding appropriateness can be discussed during blind review of data.

Have used both pre-specified and ad hoc approaches (due to recruitment issues).

Not stratified by centre:

We either include as a stratification factor (small number of centres, large patient numbers) or by including centre/provider as a cluster.

Treatment provider stratified:

Thinking about complex intervention studies, we don’t usually allow for a “provider” effect in the primary analyses, although not necessarily explicitly stated in protocol – many of these studies effectively have partial clustering. We’ve had recent interesting discussions regarding provider effect in such trials, with Chief Investigators strongly feeling that with standardised/manualised intervention and training, it isn’t relevant.

Any adjustment should be documented in SAP and final decision regarding appropriateness can be discussed during blind review of data.
Experience with multiple treatment providers is in oncology trials with different doctors delivering protocol treatment e.g. chemotherapy/radiotherapy. The actual treating doctor has not been recorded on the CRF hence all providers implicitly combined within a centre.

Have used both pre-specified and ad hoc approaches (due to recruitment issues).

Treatment providers combined by default – as we don’t routinely distinguish them in the analysis.