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

**A survey to establish current practice**

**Running head:** Managing clustering and learning effects

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

**Manuscript information**

Number of figures: 1

Number of colour figures: 0

Number of tables: 4

Number of words: 3368

Supplementary materials: Anonymised data available upon request

**Authors**

Mrs Elizabeth J Conroy (EJC) MSc. (0000-0003-4858-727X) [ejconroy@liverpool.ac.uk](mailto:ejconroy@liverpool.ac.uk) 1,2 \*

Prof. Jane M Blazeby (JMB) MD. (0000-0002-3354-3330) [j.m.blazeby@bristol.ac.uk](mailto:j.m.blazeby@bristol.ac.uk) 3

Dr. Girvan Burnside (GB) PhD. (0000-0001-7398-1346) g.burnside@liverpool.ac.uk 1,2

Assoc. Prof. Jonathan A Cook (JAC) PhD. (0000-0002-4156-6989) [jonathan.cook@ndorms.ox.ac.uk](mailto:jonathan.cook@ndorms.ox.ac.uk) 3

Prof. Carrol Gamble (CG) PhD. (0000-0002-3021-1955) [carrolp@liverpool.ac.uk](mailto:carrolp@liverpool.ac.uk) 1, 2, 4

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

2 Liverpool Clinical Trials Centre, University of Liverpool, Liverpool, UK.

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

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

5 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

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

**ABSTRACT**

**Background**

Patient outcomes can depend on the treating centre, or health professional, delivering the intervention. Health professional’s skill in delivery improves with experience, meaning that outcomes may be associated with learning. Considering differences in intervention delivery at trial design will ensure that any appropriate adjustments can be made during analysis. This work aimed to establish practice for the allowance of clustering and learning effects in the design and analysis of randomised multicentre trials.

**Methods**

A survey that drew upon quotes from existing guidelines, references to relevant publications, and example trial scenarios was delivered. Registered UK Clinical Research Collaborative registered Clinical Trials Units were invited to participate.

**Results**

Forty-four Units participated (N=50). Clustering was managed through design by stratification, more commonly by centre than treatment provider. Managing learning by design through defining a minimum expertise level for treatment provider was common (89%). One-third reported experience of expertise-based designs. The majority of Units had adjusted for clustering during analysis, although approaches varied. Analysis of learning was rarely performed for the main analysis (n=1), although was explored by other means. Insight behind the approaches used within and reasons for, or against, alternative approaches were provided.

**Conclusions**

Widespread awareness of challenges in designing and analysing multicentre trials is identified. Approaches used, and opinions on these, vary both across Units and within, indicating that approaches are dependent on the type of trial. Agreeing principles to guide trial design and analysis across a range of realistic clinical scenarios should be considered.

**MANUSCRIPT**

**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. While treatment providers are often thought of as health professionals, such as general practitioners, nurses, surgeons or therapists, the potential for clustering is also present 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). [4] 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. [2, 4, 7-10] 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. [5]

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 [11]. 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). [11] 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 [11] on the 4th January 2018 (n=51, of which 50 were registered at time of survey, *Supplementary Box 1*). As the survey involved professionals and involved discussions of current practice, no formal ethical approval were deemed necessary.

**Box 1: Example trial scenarios**

|  |  |
| --- | --- |
| **Scenarios** | |
|  |  |
| A | A trial with a large1 sample size, recruiting in several centres each with multiple treatment providers. |
|  |  |
| B | A trial with a small2 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 per site.

*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. In order to contextualise the survey content, 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. In the case that multiple responders from a single Unit provided contradictory answers, for example one responder stated they had experience and another stated they did not, it was assumed that the Unit had the experience. 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.

Free text answers were used to contextualise and illuminate quantitative responses.

**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).

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 that 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%, *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%, *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%, *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 timeto 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. While acknowledging that different approaches may be more suitable to different trial types, 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. concerns over relevance to research question, or provider not known pre-randomisation. Whilst in some settings, such as emergency treatment, advance knowledge of the treatment provider will be unobtainable. Advanced planning may be possible in other settings, such as group therapy, with guidance for practical issues like these available. [12] 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.

Managing 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, 13] Concerns were raised that identifying evidence of learning may lead to “awkward additional data summaries”.

Guidance on trial design and analysis does exist, with the most relevant of these being explicitly incorporated in to the survey questions. [4, 6, *Supplementary Box 2*] Additional documents within the ICH Series provide further guidance beyond ICH E9. [14, 15] The CONSORT statement, and relevant extensions, provide direction valuable at study design despite the document being developed to support reporting. [16, 17] The decision to explore effects may, in part, be related to the intention of the research in terms of how the results will be used and the PRECIS-2 tool has been developed to help with this. [18] However, the ability to identify and explore heterogeneity at the analysis stage is an important consideration for generalisability for all trials.

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 required by the UKCRC CTU registration process. [19] 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 trials that the CTUs conduct. Not all trials include interventions where there is learning, indeed, one would anticipate that many 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, one responder per Unit may result in experiences reported for larger Units not being indicative of all studies ran. 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, 16-17, 20] Intervention specific guidelines may have led to the variation and justifications identified in this survey. Results highlight the need for better consistency between triallists. Agreeing principles to guide trial design and analysis across a range of realistic clinical scenarios should be considered and/or further research to establish optimal methods.

**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**

CG, GB and EJC are members of the Liverpool Clinical Trials Centre, JMB is a member of the Bristol Clinical Trials and Evaluation Unit and the Bristol Randomised Trials Collaboration. JAC is a member of Oxford Clinical Trials Research Unit. No authors completed the survey on behalf of their own Unit.

**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. Trials2012**13**: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 web site.] 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 web site.] 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] | Biggs K, Hind D, Gossage-Worrall R, et al. Challenges in the design, planning and implementation of trials evaluating group interventions. *Trials*. 2020; 21, 116. |
| [13] | 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. |
| [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. |
| [15] | 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. |
| [16] | 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. |
| [17] | 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. |
| [18] | Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350:b2147. |
| [19] | 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. |
| [20] | 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. |

**ADDITIONAL TABLES:** Larger than A4

**Table 1: Methods to managing clustering and learning by design**

|  | | |  | Response statistics | | |
| --- | --- | --- | --- | --- | --- | --- |
| Question | | | Category | n | N | n/N% |
| 2 | Does your Unit have any multicentre trials that do not stratify randomisation by centre? | | Yes | 25 | 44 | 57% |
|  | No | 18 | 44 | 41% |
|  | *See Table 3 for further details.* | | No response | 1 | 44 | 2% |
| 3 | In each of the following scenarios, how was the randomisation stratified in trials that your Unit has run? Select all that apply.  *See Supplementary Table 5 for further details.* | | |  |  |  |
|  | a | Large sample size, (1) recruiting in several centres, each with multiple treatment providers | Experience in trial type | 39 | 44 | 89% |
|  |  | Centre | 34 | 39 | 87% |
|  |  | Treatment provider | 3 | 39 | 8% |
|  |  | Both | 10 | 39 | 26% |
|  |  | Neither | 1 | 39 | 3% |
|  |  | No experience in trial type | 4 | 44 | 9% |
|  |  |  | No response | 1 | 44 | 2% |
|  | b | Small sample size, (2) recruiting in several centres, each with multiple treatment providers | Experience in trial type | 31 | 44 | 70% |
|  |  | 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 | 16 | 44 | 36% |
|  |  | 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% |
|  |  |  | 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% |
|  |  |  | 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% |
|  |  | | 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% |
|  |  | | Surgeon manuals signed off by ‘senior’ surgeon prior to participation | 1 | 44 | 2% |
|  |  | | Surgical team led by consultant, who submits video measured for quality assurance, prior to participation. | 1 | 44 | 2% |
|  |  | | 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% |
| 5 | Has your Unit conducted trials with an expertise-based design, in which participating treatment providers provide only the intervention to which they have expertise? | | Yes, when applicable(3) | 13 | 44 | 30% |
|  | No, with justification | 1 | 44 | 2% |
|  | 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**

|  | | | |  | Response statistics | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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.  *See Table 4 for further details.* | | 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% |
|  | 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% |
|  |  | *See Table 4 for further details.* | | Other grouping rule or further details provided | 5 | 16 | 31% |
|  |  |  | | No | 14 | 44 | 32% |
|  |  |  | | No experience of trials of this type | 1 | 44 | 2% |
|  |  |  | | No response | 13 | 44 | 30% |
| 7 | Does your Unit include centre in the statistical model when comparing treatment? | | | Yes | 39 | 44 | 89% |
|  | But only if it was used to stratify randomisation | 18 | 39 | 46% |
|  | Always | 6 | 39 | 15% |
|  |  | | | Sometimes(1) | 15 | 39 | 38% |
|  |  | | | No, never | 3 | 44 | 7% |
|  |  | | | No response (2) | 2 | 44 | 5% |
|  | a | | If yes, and assuming that the sample size allows either, would you treat this effect as fixed or random?  *See Supplementary Box 3 for further details.* | Fixed or random, depending on circumstances | 14 | 39 | 36% |
|  |  | | Fixed | 11 | 39 | 28% |
|  |  | | Random | 12 | 39 | 31% |
|  |  | | No response | 2 | 39 | 5% |
| 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% |
|  | *See Supplementary Box 4 for further details.* | | | 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% |
|  | 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.  *See Supplementary Table 6 for further details.* | | | |  |  |  |
|  | a | Large sample size, (5) recruiting in several centres, each with multiple treatment providers | | Experience in trial type | 35 | 44 | 80% |
|  |  | 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% |
|  | 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% |
|  | 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% |
|  |  |  | | No response | 4 | 44 | 9% |
|  | In scenarios where Unit has experience, approaches to stratification changes across scenario i.e. within Unit variation to stratification | | | Different approaches across scenarios | 12 | 44 | 27% |
|  | Same approach across all scenarios | 24 | 44 | 55% |
|  | 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.* | | Yes | 32 | 44 | 73% |
|  |  | No | 9 | 44 | 20% |
|  |  | No response | 3 | 44 | 7% |
|  |  | 1. If yes to a, do you explore by graphical display? | | Yes | 31 | 32 | 97% |
|  |  | No | 0 | 32 | 3% |
|  |  | No response | 1 | 32 | 3% |
|  |  | 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?  *See Supplementary Table 8 for further details.* | | Yes | 12 | 44 | 27% |
|  |  | No | 23 | 44 | 52% |
|  |  | No response | 9 | 44 | 20% |
|  |  | i. 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% |
|  |  |  | | 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. |
| ID8 | Different statisticians/trials do different things. Often site=fixed effect and course within site = random effect. If too few within site then would combine. |
| ID14 | Retain structure at analysis. |
| ID15 | Have grouped by region / country where numbers are small. Any adjustment should be documented in the Statistical Analysis Plan and final decision regarding appropriateness can be discussed during blind review of data. |
| ID30 | Have used both pre-specified and ad hoc approaches (due to recruitment issues). |
| *Not stratified by centre:* | |
| ID32 | 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:* | |
| ID7 | 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. |
| ID15 | Any adjustment should be documented in the Statistical Analysis Plan and final decision regarding appropriateness can be discussed during blind review of data. |
| ID24 | 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 Case Report Forms hence all providers implicitly combined within a centre. |
| ID30 | Have used both pre-specified and ad hoc approaches (due to recruitment issues). |
| ID39 | Treatment providers combined by default – as we don’t routinely distinguish them in the analysis. |

**SUPPLEMENTARY MATERIAL**

**Supplementary Box 1: List of UK Clinical Research Collaborative Registered Clinical Trials Units**

The following list of 51 Units was obtained on 4th January 2019:

Barts and the London Pragmatic CTU

Barts Clinical Trials Unit

Birmingham Clinical Trials Unit

Bristol Clinical Trials and Evaluation Unit

Bristol Randomised Trials Collaboration

CaCTUS (Cancer Clinical Trials Unit Scotland)

Cambridge Clinical Trials Unit (CCTU)

Cancer Research UK Clinical Trials Unit (CRCTU)

Centre for Healthcare Randomised Trials (CHaRT)

Centre for Trials Research

Comprehensive CTU @ UCL

CR UK & UCL Cancer Trials Centre

Diabetes Trials Unit (Churchill Hospital, Oxford)

Edinburgh Clinical Trials Unit, Edinburgh

Glasgow Clinical Trials Unit

Imperial Clinical Trials Unit

Intensive Care National Audit & Research Centre (ICNARC) CTU

Keele Clinical Trials Unit

King's Clinical Trials Unit at King's Health Partners

Leeds Clinical Trials Research Unit

Leicester Clinical Trials Unit

Liverpool Trials Collaborative

London School of Hygiene & Tropical Medicine

Manchester Academic Health Science Centre Clinical Trials Unit (MAHSC-CTU)

Medical Research Council Clinical Trials Unit at UCL

Newcastle Clinical Trials Unit (NCTU)

NHS Blood and Transplant Clinical Trials Unit

North Wales Organisation for Randomised Trials in Health (NWORTH)

Northern Ireland Clinical Trials Unit

Norwich Clinical Trials Unit

Nottingham Clinical Trials Unit

NPEU Clinical Trials Unit

Oxford Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU)

Oxford Clinical Trials Research Unit (OCTRU)

Oxford Primary Care and Vaccines Collaborative Clinical Trials Unit

Papworth Trials Unit Collaboration

Peninsula Clinical Trials Unit

PRIMENT Clinical Trials Unit at UCL

Royal Marsden Clinical Trials Unit (RM-CTU)

Sheffield Clinical Trials Research Unit

Southampton Clinical Trials Unit

Surrey Clinical Trials Unit

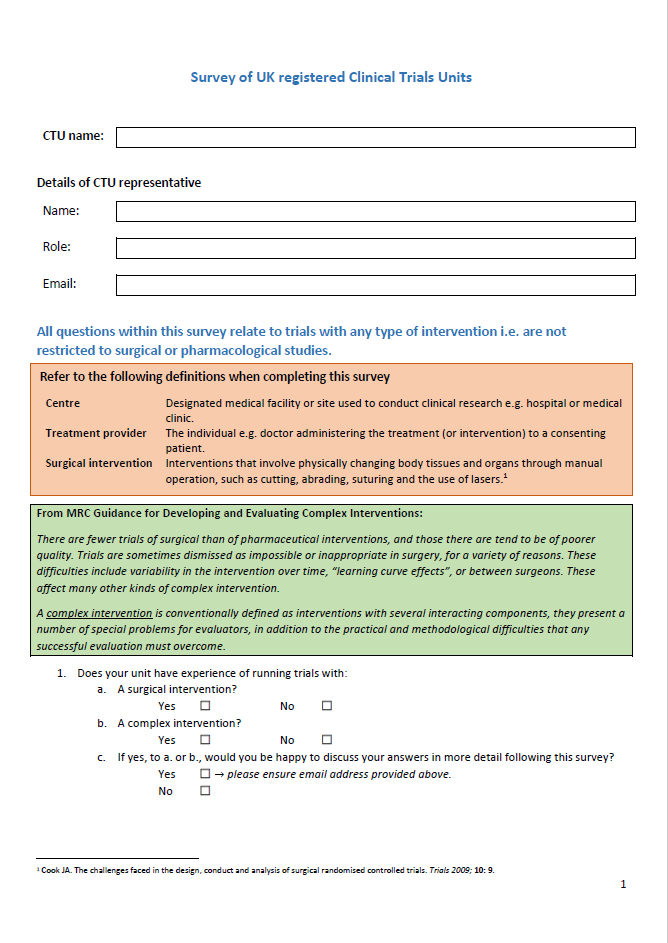
Swansea Trials Unit

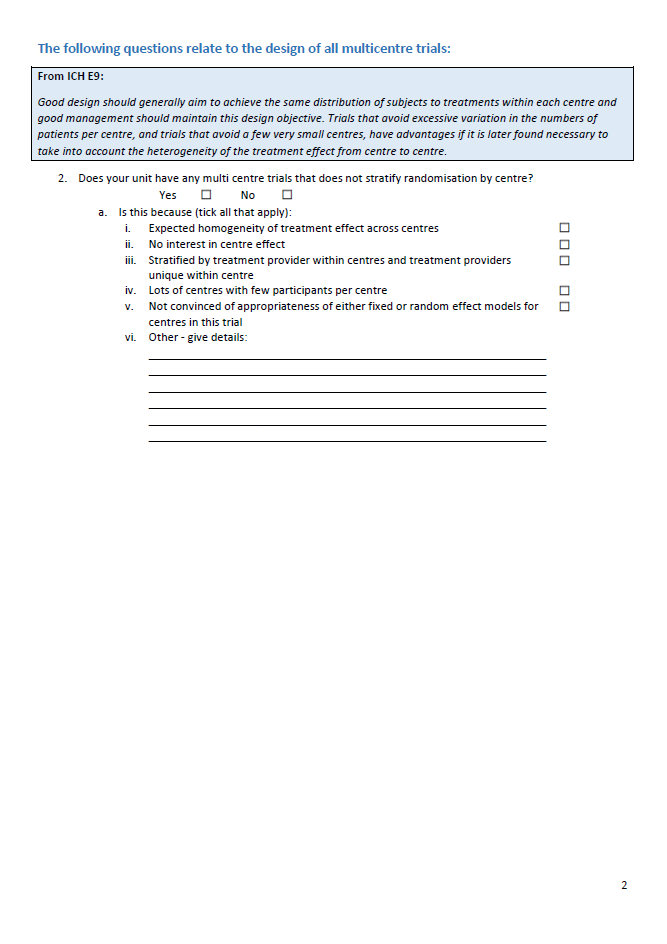
Tayside Clinical Trials Unit

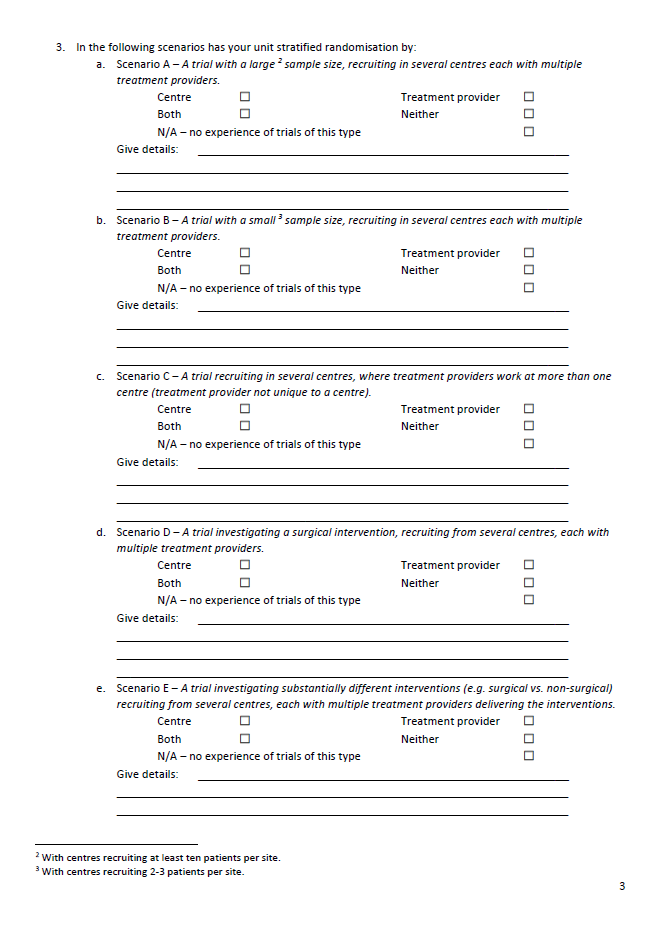
The Institute of Cancer Research Clinical Trials & Statistics Unit (ICR- CTSU)

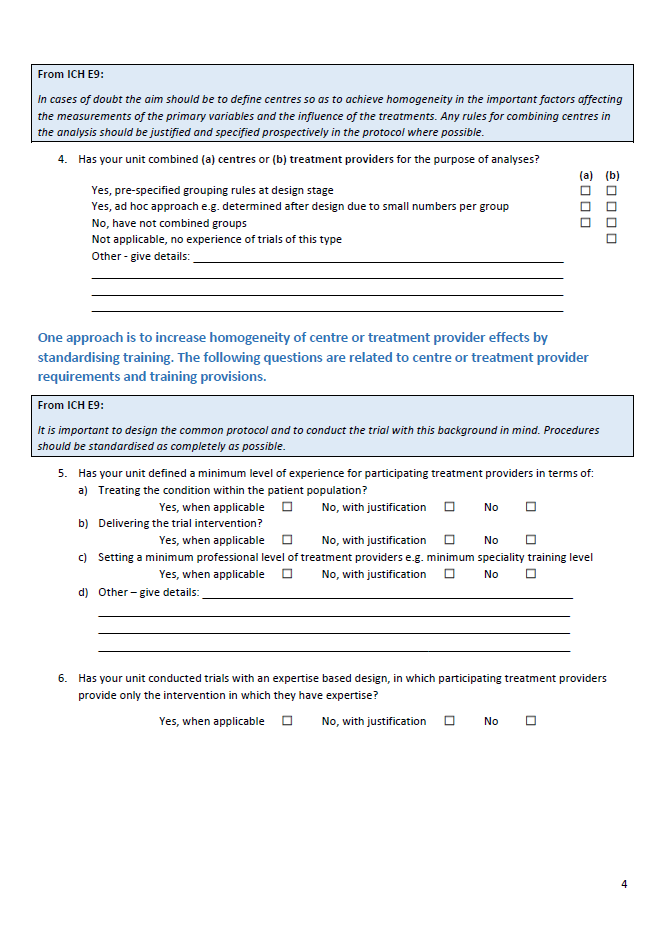
Warwick Clinical Trials Unit

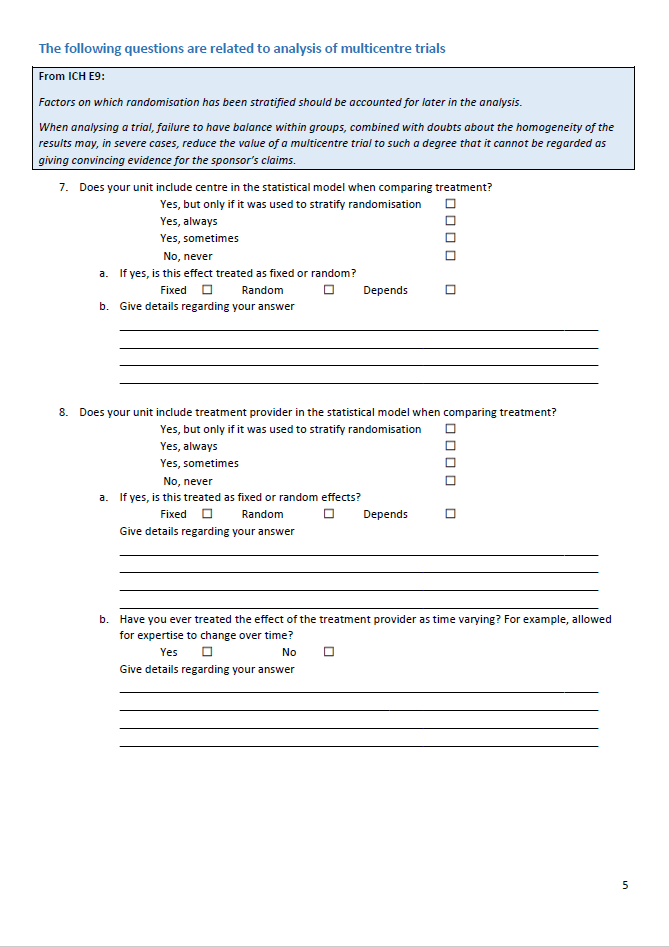
York Trials Unit

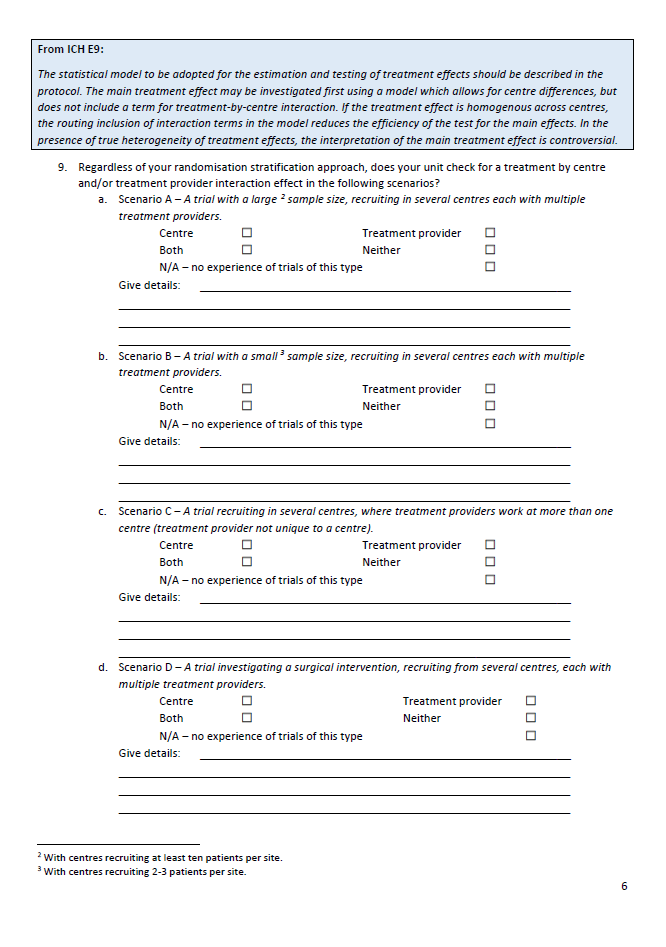
**Supplementary Box 2: Survey**

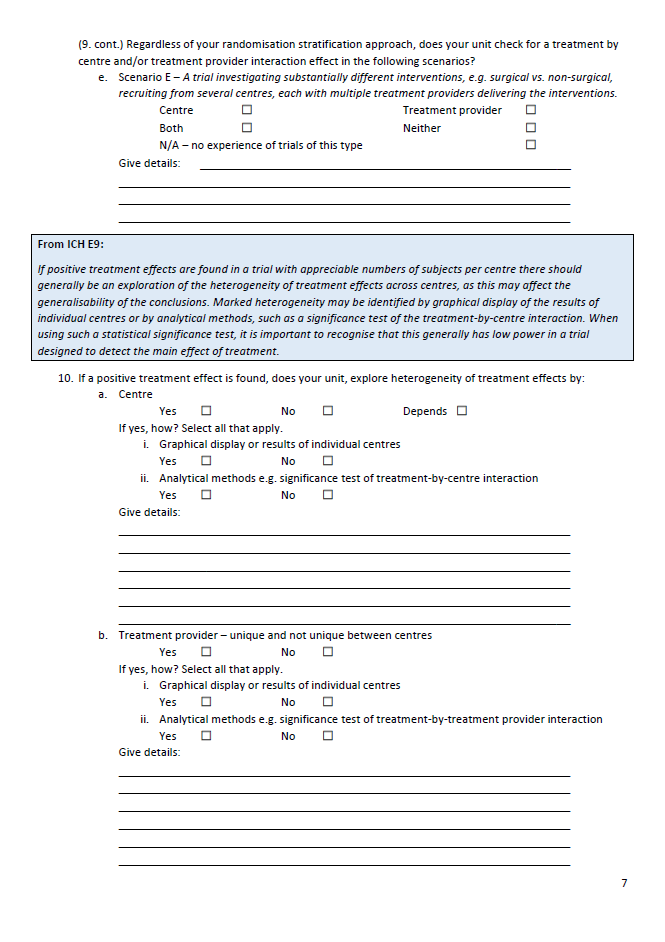


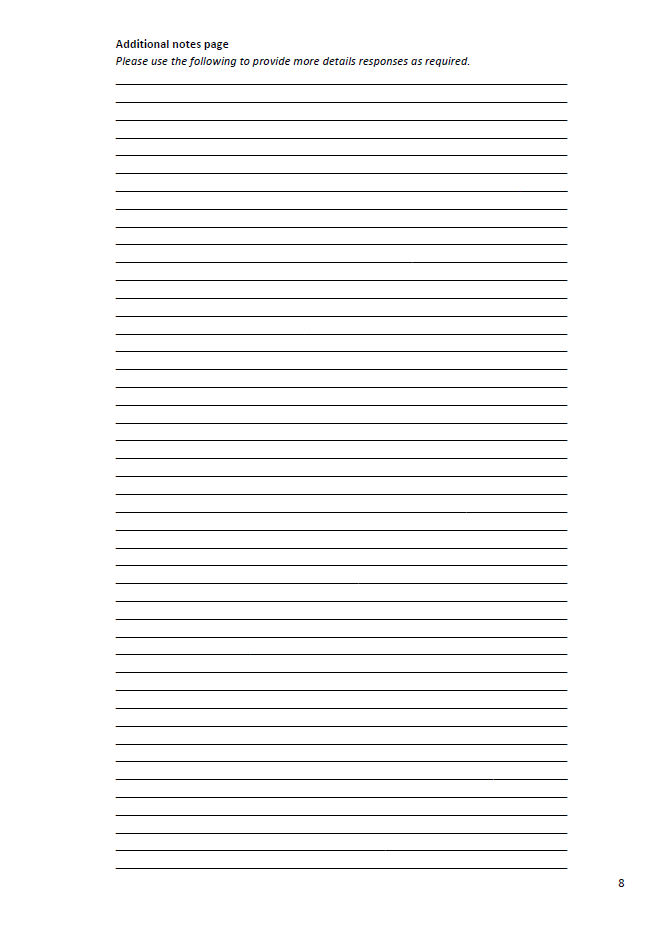












**Supplementary Table 1: CTU completion rate**

|  |  |  |  |
| --- | --- | --- | --- |
| Participation status | Number of Units | | |
|  |
|  | n | N | n/N% |
| Completed | 44 | 50 | 88% |
| Pen and paper | 31 | 44 | 70% |
| Electronic form | 13 | 44 | 30% |
| Declined, unable to participate | 1 | 50 | 2% |
| Declined, no reason provided | 5 | 50 | 10% |

**Supplementary Table 2: Role of Unit representative**

|  |  |  |  |
| --- | --- | --- | --- |
| Role | Number of Units | | |
|  |
|  | n | N | n/N% |
| Professor/Reader of Medical Statistics | 3 | 44 | 7% |
| Director of Unit | 6 | 44 | 14% |
| Statistics lead/Head of statistics | 13 | 44 | 30% |
| Senior statistician | 15 | 44 | 34% |
| Statistician | 7 | 44 | 16% |

**Supplementary Table 3: CTU completion rates by speciality trial types**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Specialises in trial of type | | | | | | | | | | | |
|  |  | Complex intervention | | | | Surgical intervention | | | | Cluster randomised | | | |
| Completer status | | Yes  N=39 | | No  N=11 | | Yes  N=36 | | No  N=14 | | Yes  N=17 | | No  N=33 | |
| N | n/N% | N | n/N% | n | n/N% | n | n/N% | n | n/N% | n | n/N% |
| Completed | | 35 | 90% | 9 | 82% | 33 | 92% | 11 | 79% | 16 | 94% | 28 | 85% |
| Declined, unable to participate | | 1 | 3% | 0 | 0% | 1 | 3% | 0 | 0% | 1 | 6% | 0 | 0% |
| Declined, no reason provided | | 3 | 8% | 2 | 18% | 2 | 6% | 3 | 21% | 0 | 0% | 5 | 15% |

**Supplementary Table 4: Experience in running complex and/or surgical interventions**

|  | | |  | Response statistics | | |
| --- | --- | --- | --- | --- | --- | --- |
| Question | | | Category | n | N | n/N% |
| 1 | Which of the following intervention trials does your Unit have experience of running | | Both surgical and complex interventions | 25 | 44 | 57% |
|  | Surgical interventions only | 4 | 44 | 9% |
|  | Complex interventions only | 7 | 44 | 16% |
|  |  |  | Neither | 7 | 44 | 16% |
|  |  |  | No response | 1 | 44 | 2% |

**Supplementary Table 5: Comments on stratification approaches in example scenarios (Question 3)**

| ID | Scenario | Stratification approaches | Free text |
| --- | --- | --- | --- |
| ID1 | A | No experience | For A, our trials are in a critical care setting – this is a team based specialty and it is generally not possible to distinguish individual treatment providers |
|  | B | No experience |  |
|  | C | No experience |  |
| ID3 | A | No experience | Have not run studies with several centres. A study with two sites that we run stratifies randomisation by neither centre nor treatment provider. |
|  | B | No experience |
|  | C | No experience |
| ID4 | A | Centre only  Both | For A, most commonly just by centre |
|  | B | No experience |  |
|  | C | Centre only |  |
| ID8 | A | Centre only | I do not think we have any behavioural therapy/course type interventions at our Unit, but stratifying by site in these seems OK. If you were to stratifying by deliverer, there is likely to be too few per strata. For cross nesting, this is often delivered by >1 treatment provider. There is a group effect or course effect rather than therapy effect in this case. For example, personalities or dynamics. I do not see this changing unless convinced otherwise. |
| B | Centre only |
| C | Centre only |
| D | Centre only | Responses based on one trial. Not always known who surgeon will be in advance and too few per strata as often one surgeon. But here, I could be convinced. Note: In previous trials that I have analysed (1 surgery, 2 group) observed ICC=0. Again, anecdotal but I am unconvinced we need to change. |
| E | No experience |
| ID10 | A | Centre only | For A, centre: ambulance station. For B, feasibility: centre: ambulance station – provider: paramedic. |
|  | B | Centre only  Both |  |
|  | C | No experience |  |
|  | D | Centre only  Treatment provider only  Both | For D, centre: Hospital; Treatment provider: operating surgeon, closing the wound. |
|  | E | No experience |  |
| ID13 | A | Centre only | We rarely collect data on treatment provider. |
| B | Centre only |
| C | No experience |
| ID14 | A | Centre only  Both | For a, dependent on trial. General comments - considered and decided on a trial by trial basis. Don't feel you can always standardise. |
|  | B | Centre only |
|  | C | Treatment provider only |  |
|  | D | Centre only  Treatment provider only | General comments - considered and decided on a trial by trial basis. Don't feel you can always standardise. |
|  | E | Centre only |
| ID15 | A | Centre only  Treatment provider only  Both | We would generally stratify by centre as a rule of thumb if we’re happy we'll get enough patients per site. Stratifying by treatment provider is only carried out currently on the larger surgical studies = again where we're confident that there will be enough patients. Also depends on what other stratification factors are needed. |
|  | B | Centre only |
|  | C | Centre only |
| ID17 | A | Centre only | It is assumed that differences in treatment will mainly be due to differences in facilities in each centre and different treatment protocols within each centre. |
| B | Centre only |
| C | Centre only |
| ID18 | A | Centre only  Neither centre nor treatment provider | Recent conversions between senior statisticians advocate not stratifying by centre in any situation. They cited concerns regarding prediction of allocation. |
| B | Neither centre nor treatment provider |
| C | No experience |
| D | No experience | For D and E, not aware of any locally, but if we did then think definitely by treatment provider. |
| E | No experience |
| ID29 | A | Both | For C, stratification is likely to be chosen by the expected homogeneity, so may be centre or treatment provider. This will be intervention specific. |
| B | Centre only |
| C | Centre only  Treatment provider only |
| ID30 | A | Centre only | Only centre used as treatment provider could vary during the trial. This would add logistics of then having to update the randomisation protocol. |
| B | No experience |
| C | No experience |
| ID32 | A | Centre only | We often include treatment providers as a cluster effect but do not usually stratify as do not always know at randomisation. For c specifically, centres combined out of necessity. |
| B | Centre only |
| C | Centre only |
| D | Centre only | For D and E, usually comparing the intervention policy not the different aspects of the intervention. We try to standardise the interventions to make them as similar as possible. We have done surgery vs. physiotherapy for example. |
| E | Centre only |
| ID35 | A | Centre only | I can't remember every detail of every study on our books, I've written down my best guess. I don't know detail for other statistician's studies. |
|  | B | Neither centre nor treatment provider |
|  | C | Neither centre nor treatment provider |
| ID39 | A | Centre only | For A, Surgeons within seven centres. Surgery is conducted by a team which includes a variable subset of surgeons within a centre. Hence stratified by centre only, as would be unclear which surgeons to stratify by. For B, several pilot trials like this – to much stratification with a small sample size may not achieve balance across trial groups. |
|  | B | Centre only |
|  | C | No experience |
|  | D | Centre only | For D, Surgeons within seven centres. Surgery is conducted by a team which includes a variable subset of surgeons within a centre. Hence stratified by centre only, as would be unclear which surgeons to stratify by. For E, Surgery versus radiotherapy versus nurse-led active monitoring. Difficult to see how strata for randomisation could be defined at the practitioner level. |
|  | E | Centre only |

**Supplementary Box 3: Reasons for using fixed or random effect for centre (Question 7)**

|  |  |
| --- | --- |
| *Use both fixed effect or random effect, as required (n=14):* | |
| ID1 | Always included in adjusted analyses, but sometimes a simple unadjusted analysis is also presented with no accounting for centre. Usually included as a random effect, except in studies with only a small number of centres e.g. pilot studies. |
| ID6 | Random if many centres, fixed if few. Also, some types of model don’t allow random effects e.g. quantile regression so no choice but to use fixed effects. |
| ID7 | If centre was used for stratification or minimisation then included in the primary statistical model. If centre wasn’t a stratification factor, on some occasions, centre might be included in exploratory/sensitivity analyses – partly depending on how many centres and sample size. The choice between fixed and random effects is trial specific – partly driven by whether the centres can be considered and/or justified as being a random sample and also the number of centres (low number would usually be modelled as fixed effects). |
| ID10 | Choice of fixed effects or random effects depends on number and/or nature of centres. |
| ID12 | Mainly random unless small number of centres. |
| ID14 | When a small number of centres, use fixed. |
| ID30 | Have treated as both but more likely to be random as this is more reflective of what we need. |
| ID35 | Random is >5 centres, fixed is lower number. (I think! Can’t remember the precise numeric cut off). Can’t do random effects very well if the number of centres is too low. |
| ID38 | Usually random, but if few centres (5 or less) I would use a fixed effect for centre. |
| ID39 | Centre is almost always included in the model – other than when a lot of centres have recruited only one or two patients e.g. primary care trials. Tends to be included as a fixed effect in a standard multi-centre trial, and a random effect in cluster randomised studies. |
| *Fixed effect for centre (n=11):* | |
| ID2 | Based on limited experience, centre was specified in the protocol as a fixed effect for one trial. |
| ID8 | Usually! I think it is easier to assess treatment by strata interactions in fixed effect. Debate is inevitable (even Stephen Senn says no firm views on this). |
| ID20 | Only one multivariate trial with four large centres and low heterogeneity between centres in the treatment effect. |
| ID27 | It requires fewer assumptions and easier to explain. If exact balance is achieved then the maths will give identical estimated standard errors with both models. |
| *Random effect for centre (n=12):* | |
| ID9 | We are interested in the impact of population of centres rather than the sample of centres used in the trial. |
| ID11 | Would aim for random effects but if not feasible (too few centres or too few per centre) then would include as a fixed effect. |
| ID13 | Uses fewer degrees of freedom if many centres and that we do not expect structured differences between centres. |
| ID15 | Usually an underlying assumption that centre may be a surrogate for socio-economic factors that may affect outcome and/or treatment effect so often not happy to assume that there is an equal fixed treatment effect across all sites. Would use random due to large number of centres. |
| ID23 | Depends a bit on the number of centres. Need to have a sufficient number to preserve the degrees of freedom. |
| ID29 | To report the centres as a sample of centres. |
| ID32 | Varies between statisticians – going more down the random effect moving forwards. |

**Supplementary Box 4: Reasons for using fixed, random or time varying effect for treatment provider (Question 8)**

|  |  |
| --- | --- |
| *Use both fixed effect or random effect, as required (n=4):* | |
| ID7 | If treatment providers was used for stratification or minimisation then included in primary statistical model. For most of our complex intervention trials, we wouldn’t know who the provider is going to be at the time of randomisation and so can’t be a stratification/minimisation variable. We’ve considered e.g. partial clustering at the design stage where the intervention group has providers and the control group has treatment as usual, but usually insufficient information available, leading to potential for secondary/exploratory analyses of treatment provider. |
| ID35 | Depends on the number of providers. Can’t do random effects if number of providers is too low. Random is >5 centres and fixed is lower. I can’t think of an example where we have adjusted but we would if it were sensible. |
| *Fixed effect for treatment provider (n=2):* | |
| ID27 | It requires fewer assumptions and easier to explain. If exact balance is achieved then the maths will give identical estimated standard errors with both models. However, never actually done this. |
| ID30 | This is dependent on the data and requirements. |
| *Random effect for treatment provider (n=16):* | |
| ID8 | There are too few per strata in our trial to consider as fixed. May be qualitative assessment. |
| ID10 | Choice of random effect is based on parsimony. |
| ID15 | If treatment provider is included as a stratification factor, it’ll be because we’re concerned the provider will have an impact on the outcome but also because we’d expect different populations for different treatment providers. We have no interest in therapist effect for main adjustments, we would adjust to complement randomisation stratification factor. |
| ID23 | Depends on number, need to preserve degrees of freedom. |
| ID29 | To represent the result as a sample. |
| ID32 | As a cluster random effect. |
| ID39 | When a clear provider structure to the data, we will pre-specify the main analysis to accommodate this. We would try and capture the hierarchical structure in a multi-level model. |
| *Time-varying effect used for treatment provider (n=2):* | |
| ID38 | 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!) |
| *Time-varying effect not used for treatment provider (n=19):* | |
| ID7 | Interesting idea! Thinking again about our complex intervention trials, I’m not entirely sure that we would expect to see a change in treatment effect over time, as the Chief Investigator would, I suspect, say that with manualised interventions, etc, this shouldn’t be observed. But something to think about…. |
| ID8 | Secondary exploratory analyses. Had we found evidence of learning we would have had awkward additional data summaries/presentations. |
| ID10 | Not yet appropriate. |
| ID15 | Not yet appropriate in trials that we run – plan to discuss exploration of learning curve of treatment provider in one of our ongoing studies. |
| ID23 | Did not consider. |
| ID30 | Often time restrictions on completing. |
| ID32 | Unusually doesn’t change during the treatment period. |
| ID35 | I did a sensitivity analysis once to check for learning effects, but there was no evidence of it. I don’t know what other senior statisticians at our Unit have done. |
| ID39 | Not in the main analysis – aim for practitioner skill to have stabilised. |

**Supplementary Table 6: Comments on interaction investigation in example scenarios (Question 9)**

| ID | Scenario | Stratification approaches | Free text |
| --- | --- | --- | --- |
| ID2 | A | No experience | Mainly worked on non-randomised trials in my Unit and so none of the above are applicable. |
|  | B | No experience |
|  | C | No experience |
| ID4 | A | Centre only  Neither centre nor treatment provider | For A, likely to investigate interaction for any stratification factors – so may include centre but, may be geographical region instead. |
|  | B | No experience |  |
|  | C | Centre only  Neither centre nor treatment provider |  |
| ID5 | A | Neither centre nor treatment provider | For A, not formally. For C, not formally. |
|  | B | Neither centre nor treatment provider |  |
|  | C | Neither centre nor treatment provider |  |
| ID6 | A | No response | For A, sometimes, specific to each trial. Depends on the nature of the intervention and numbers within treatment provider. For C, depends on numbers and nature of intervention |
|  | B | Neither centre nor treatment provider |  |
|  | C | No response |  |
|  | D | No response | For D, depends on numbers and nature of intervention. For e, depends on numbers and nature of intervention. |
|  | E | No response |  |
| ID7 | A | Centre only  Treatment provider only | For A, would usually (given sufficient numbers) look at centre\*treatment group i/a; on a limited number of occasions looked at provider\*treatment group i/a |
|  | B | Neither centre nor treatment provider |  |
|  | C | No experience |  |
| ID8 | A | Both centre and treatment provider | Some investigation is often useful - maybe qualitative. Rare to detect significant, clear interaction and therefore often summarise descriptively. |
| B | Centre only |
| C | Centre only |
|  | D | Both centre and treatment provider | Only have experience of one trials in Scenario D. We did an informal assessment for individual surgeon. |
|  | E | No experience |
| ID9 | A | Centre only | Centres are now nested within the treatment provider and makes a multilevel model. |
| B | Centre only |
| C | No experience |
| ID10 | D | No experience | For D, only one such trial in my experience – data collection and analysis plans still in development. |
|  | E | No experience |  |
| ID13 | A | Neither centre nor treatment provider | Has never been of interest. |
| B | Neither centre nor treatment provider |
| C | No experience |
| ID14 | A | Centre only  Treatment provider only | For A, maybe exploratory as usually not powered. For B, maybe exploratory as usually not powered. General comments: Trial by trial decision as appropriate in relation to interpretation and associated power. |
|  | B | Neither centre nor treatment provider |  |
|  | C | Treatment provider only |  |
|  | D | Centre only  Treatment provider only | For D, if specialist. For E, if specialist. General comments: Considered and decided at trial by trial basis - depends on phase and question. |
|  | E | Centre only  Treatment provider only |  |
| ID15 | A | Centre only  Treatment provider only | Might group in 'small' trials and do a treatment\*county interaction for example. Often do subgroup analysis and plot treatment effect within site in a forest plot. |
|  | B | Neither centre nor treatment provider |  |
|  | C | Neither centre nor treatment provider |  |
| ID22 | A | Neither centre nor treatment provider | For A, most trials we have adjusted for the centre and treatment provider but not performed any treatment-by-centre interaction. |
|  | B | Neither centre nor treatment provider |
|  | C | Neither centre nor treatment provider |
| ID30 | D | Treatment provider only | Only limited experience in our Unit of d. |
| E | No experience |
| ID32 | A | Centre only | We usually look at centre by treatment interaction to assess whether the treatment effect is similar across centres but do not normally present this as part of the model. |
| B | Centre only |
| C | Centre only |
|  | D | Centre only | We usually look at centre by treatment interaction to assess whether the treatment effect is similar across centres but do not normally present this as part of the model. |
|  | E | Centre only |
| ID35 | A | Neither centre nor treatment provider | Providing best guess here. |
|  | B | Neither centre nor treatment provider |
|  | C | No experience |
| ID38 | A | Neither centre nor treatment provider | For A, I’ve put neither because we don’t routinely do it. I do usually work up a forest plot to explore within centre treatment effects, no formal treatment-by-centre interactions. |
|  | B | Neither centre nor treatment provider |
|  | C | No experience |
|  | D | Neither centre nor treatment provider | For D, again I’ve put neither because we don’t routinely do it. I do usually work up a forest plot to explore within centre treatment effects, no formal treatment-by-centre interactions, but I can’t recall drilling down to surgeon level in forest plots. |
|  | E | Neither centre nor treatment provider |
| ID39 | A | Neither centre nor treatment provider | For A, not done routinely – if interest in a centre effect in a particular trial, an investigation of the treatment by centre interaction would be pre-specified. For B, not done routinely, and would point out the problems of looking at interaction effects when sample size is small if the chief investigator requested such an analysis. |
|  | B | Neither centre nor treatment provider |
|  | C | No experience |
|  | D | Neither centre nor treatment provider | For D, not done routinely – if interest in a centre effect in a particular trial, an investigation of the treatment by centre interaction would be pre-specified. For E, not done routinely – if interest in a centre effect in a particular trial, an investigation of the treatment by centre interaction would be pre-specified. |
|  | E | Neither centre nor treatment provider |

**Supplementary Table 7: Details of how Units explore heterogeneity by centre in the presence of a treatment effect (Question 10A)**

| ID | Graphical display | Analytical methods | Further details on exploring heterogeneity by centre |
| --- | --- | --- | --- |
| ID1 | Yes | Yes | Depends |
| ID2 | Yes | Yes | Interaction tests done where necessary. |
| ID3 | Yes | No | We generally work with small studies so far and significance testing of treatment by centre would not seem appropriate or reliable with such small samples. |
| ID4 | No | Yes | Depends |
| ID6 | Yes | Yes | Depends |
| ID7 | Yes | Yes | Depends on whether any difference between centres is likely to be relevant of not. We would usually present some basic graphs and summary statistics at the centre level; sometimes pre-specified centre\*treatment will be assessed, but always with the caveat of likely to have low power. In larger trials, sometimes presented Chief Investigators for the treatment effect by centre. |
| ID8 | Yes | Yes | Happened once only and we'd already planned to do this. This was a stepwise trial - hoping someone will publish at some point. Wanted to do this for fidelity assessment - qualitative. |
| ID10 | Yes | Yes | Graphical methods usually take precedence; supported by analytical methods. |
| ID14 | Yes | Yes | Depends |
| ID15 | Yes | Yes | One trial with a positive treatment effect explored variation by centre graphically and by descriptive statistics. This was not done due to positive treatment effect and was pre-planned. Another explored through forest plots - this has post hoc and in a study where there was no overall treatment effect. |
| ID18 | Yes | Yes | Forest plot and test. |
| ID19 | Yes | Yes | Tabulations, multivariate models, etc. |
| ID21 | Yes | No | Forest plots typically. |
| ID22 | Yes | Yes | Most of the trials have explored the heterogeneity of treatment effects across centres by either graphical or analytical methods. |
| ID23 | Yes | Yes | Not significance tests. |
| ID27 | Yes | No | Depends. Must by prespecified and of interest a priori. |
| ID30 | Yes | Yes | Treatment by centre interaction testing but rarely go beyond this - requires careful consideration of the implication of explanation to the Chief Investigator. |
| ID32 | Yes | No | Don't use significance tests as no power for these. |
| ID34 | Yes | No | Explore by region. |
| ID35 | Yes | No | Don't do this routinely but we probably do it a fair amount. If we did it, we would probably try to estimate treatment effect within each centre but this wouldn't be possible for tiny centres. |
| ID39 | Yes | Yes | Not done routinely - if interest in a centre effect in a particular trial, and investigation of the treatment by centre interaction would be pre-specified. If not pre-specified, we would not allow an interaction to be prompted by a positive treatment effect. |
| ID42 | Yes | Yes | Depends |

**Supplementary Table 8: Details of how Units explore heterogeneity by treatment provider in the presence of a treatment effect (Question 10B)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Effect explored | Graphical display | Analytical methods | Further details on exploring heterogeneity by treatment provider |
| ID1 | No response | NA | NA | No experience in trials of this type. |
| ID2 | No response | NA | NA | No experience in trials of this type. |
| ID3 | No | No | No | Generally we do not consider treatment provider effects. After today, I will take this back to our Unit and see if we should start taking it into account in our designs and analyses. |
| ID5 | No response | NA | NA | No experience in trials of this type. |
| ID6 | Yes | Yes | Yes | Depends. |
| ID7 | Yes | Yes | Yes | The provider effect doesn't tend to be of particular interest in our trials, although I believe that in complex intervention trials, the provider effect can be underplayed. |
| ID8 | No | No | No | Too awkward - often we do course/group interventions with course leads swapping in/out/leaving etc. |
| ID14 | Yes | Yes | Yes | Sometimes, this is trial dependent - if relevant and interpretable. |
| ID35 | Yes | No | No | Need to be careful not to suggest some clinicians are worse than others (similarly for centres). |
| ID39 | No | No | No | Not done routinely and I don't think we have pre-specified this. |
| ID42 | Yes | Yes | Yes | Depends. |