Learning and clustering: statistical adjustment for the learning curve and clustering effects in randomised surgical trials

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by:

Elizabeth Jane Conroy

December 2021

Abstract

Learning and clustering: statistical adjustment for the learning curve and clustering effects in randomised surgical trials

Elizabeth Jane Conroy

The need for more, and better, randomised surgical trials is well recognised, and recently the number of surgical trials has grown. Rigorous design and analysis of such studies is important to support clinical decision making. Two associated methodological challenges are clustering effects, by centre and surgeon, and the surgical learning curve. This thesis aims to improve the design and analysis of surgical trials by investigating existing guidance, establishing current practice and demonstrating how design and analysis can incorporate clustering and learning.

Existing guidance for managing these challenges exist, but this work identifies that they focus more on design than analysis. As there is no single document, triallists must access multiple documents to gain full understanding, ultimately leading to inconsistencies in practice.

Two novel reviews of surgical trials are undertaken covering a time period of remarkable growth of surgical trials. The first, of 247 published trials, demonstrates that clustering and learning considerations are underreported, methods used to do so vary and reporting guidelines are poorly adhered to. It is recommended that triallists report these methods, or justify where not, to support results interpretation. Early consideration of these effects is vital. The second, comprising fifty funded grant applications by a leading UK funder, identifies early consideration of these effects and the funder as a potential driver of better practice. Recommendations are provided about when and how to address surgical learning and clustering in the design and analysis.

To complete understanding of current practice, forty-seven statisticians from UK clinical trials Units were surveyed. Widespread awareness of challenges in design and analysis are identified. Approaches used to manage clustering and learning varies both across and within Units, suggesting that agreed principles, across a range of trial scenarios, are needed. A number of real surgical trials, varying by intervention and setting, are presented as a practical demonstration of approaches to design and analysis.

Statistical methods for exploring the presence of clustering and learning, by centre and surgeon, and any impact on trial conclusions are demonstrated using real trial datasets. For clustering, simulated data were used to explore the impact of clustering under different scenarios. Clustering became a greater concern as the intraclass correlation and true treatment difference increased. For learning, a curve was identified but it did not impact trial outcomes. Developing better measures of learning for use within such explorations is recommended.

Good design can minimise the impact of clustering and learning, but statistical methods that fail to account for these effects, if present, can lead to biased treatment estimates and reduced power. Clustering and learning should be managed using a design and analysis approach. Considerations should be made early, and on a trial-by-trial basis, to ensure that the trial conclusions are valid.

Acknowledgements

For financial support, I am grateful to the National Institute for Health Research who gave me the opportunity to undertake this project by providing funding.

I would like to thank my supervisors, Carrol for her constant support, advice, patience and guidance throughout the past ten years, Girvan for answering every question without judgement, and Jane and Jonathan for their invaluable methodological insights, technical support and kind words of encouragement throughout.

I am also indebted to numerous other colleagues in my department, for their support throughout the past decade; in particular Anna Rosala-Hallas, for her infectious enthusiasm that kept me sane throughout this marathon (and the other) and Laura Bonnett, for being my companion throughout the daily highs and lows of an NIHR fellowship and the hourly highs and lows of motherhood.

I am eternally grateful for the unwavering support of my husband Dean, my mum for teaching me that *"can't means won't"*, and my son Jasper who helped me to put all of this into perspective.

Table of contents

List of tables

List of figures

Abbreviations

CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CTU	Clinical Trials Unit
Cusum	Cumulative sum
EME	Efficacy and Mechanism Evaluation
GEE	Generalised estimating equations
HTA	Health Technology Assessment
ICC	Intraclass correlation coefficient
ICH	International Council for Harmonisation
JCE	Journal of Clinical Epidemiology
Max	Maximum
Min	Minimum
NETSCC	NIHR Evaluation, Trials and Studies Coordinating Centre
NIHR	National Institute for Health Research
OR	Odds ratio
RCT	Randomised controlled trial
SD	Standard deviation
UKCRC	UK Clinical Research Collaboration
VPS	Ventriculoperitoneal shunts

Initials of researchers involved in this thesis

ARH	Anna Rosala-Hallas
CG	Carrol Gamble
EJC	Elizabeth J Conroy
GB	Girvan Burnside

JAC Jonathan A Cook

JMB Jane M Blazeby

Chapter 1 : Introduction

1.1. Learning and clustering within randomised surgical trials

Surgery is an indivisible, indispensable part of health care. (1) Surgical conditions comprise a broad range of diseases that represent approximately 30% of the global burden of disease and span 100% of disease subcategories. (1, 2) Despite the importance of surgery, there has long been a paucity of high-quality evidence within the field, with Richard Horton, Editor of the Lancet, terming surgical research as a 'comic opera' in 1996. (3) In 2012, despite one-third of UK hospital admissions involving surgery, less than 2% of government funding for medical research was invested into it. (4)

Over the past decade, the path towards major change to address these shortfalls has been taken. In 2012, following the National Health Service deeming surgery a high research priority, the National Institute for Health Research (NIHR) released a themed call for Applied Health Research in Surgery and the Royal College of Surgeons of England initiated a nationwide Surgical Trials Programme. (5, 6) In 2015, the Lancet opened a Commission on Global Surgery and the NIHR Global Health Research Unit in Global Surgery was established, to open research hubs in low- and middle-income countries across the world. (7, 8) These new initiatives and collaborations all aim to improve and increase surgical research, ultimately leading to a growth in surgical trials worldwide.

To ensure this investment is not wasted, we need to ensure the validity of the clinical decisions that these trials inform. At the time of writing, there are 252 and 189 active randomised surgical trials registered on the ClinicalTrials.gov website and funded by the NIHR respectively, a figure that is likely to underestimate the true volume of surgical research. (9, 10) With the number of trials set to increase further, and becoming more geographically

dispersed, ensuring that these trials are designed and analysed with the highest possible rigour will support clinical decision-making.

Randomised controlled trials (RCTs) are considered the highest level of evidence, second only to systematic reviews of such trials. (11) There are many practical and methodological difficulties that a medical researcher must overcome to conduct a successful RCT. In trials with a surgical intervention, these difficulties are often magnified. (12-15) Surgical interventions, delivered as an intervention or as a setting, consist of many interacting components - such as the procedure itself, surgeon expertise, and pre or postoperative care. (14)

The surgical learning curve, where a surgeon's expertise increases throughout the course of a trial, and clustering, where variation in outcomes may be smaller between patients treated by the same surgeon, surgical team or centre than those treated by different surgeons, teams or centres, are two such methodological challenges associated with randomised surgical trials. (12, 14) An unpublished review of randomised surgical trials, conducted prior to 2015, within the National Institute of Health Research Health Technology Assessment (NIHR HTA) series was undertaken when developing this research plan. In the 16 monographs reporting a randomised surgical trial, half discussed the surgical learning curve as a challenge, one-third identified clustering by centre and 13% clustering by surgeon. It is recommended that these issues are considered in multicentre trials (16) and may have increased relevance within surgical trials depending on the interventions under investigation and their levels of routine use. (12, 14, 16-18)

Learning curves and clustering have so far been investigated in isolation, often within specific fields and including studies of observational design (19, 20) When designing randomised surgical trials, it is important to consider homogeneity of the treatment effect and therefore the potential existence and impact of both learning and clustering, by centre and surgeon. This

should be done as early as possible during trial design to avoid issues arising that may violate the validity of the trial results. In extreme cases, where the trial results are questioned by the research community, the trial team should be prepared to alleviate doubts of heterogeneity of treatment effects. (16) Whilst the notion of learning and clustering in analysis is familiar to many statisticians, the extent to which, when and how these considerations are made is unknown.

1.2. Aims

The aim of this thesis is to improve the design, analysis and generalisability of randomised surgical trials. A comprehensive investigation into existing guidance for, and current practice of, management of the surgical learning curve and clustering effects, at the centre and surgeon level, during design, analysis and reporting of these trials will be undertaken. Barriers to, and drivers to support, consideration of these effects will be established. In addition, this work aims to demonstrate:

- How learning and clustering can be managed at trial design, and
- How the presence of learning and clustering can be explored and, if necessary, adjusted for within the trial analysis.

1.3. Thesis structure

This chapter serves as background to the research that is presented within this thesis. The rest of the thesis is structured as follows.

In *Chapter 2*, existing guidance developed to support the design, conduct and reporting of RCTs is identified and summarised. An overview of aspects that have relevance to the learning curve and clustering effects is provided.

Chapter 3 provides a review of published RCTs. The degree of learning and clustering considerations, and the extent to which triallists reported attempts to design and analyse the trial with these in mind, is explored. In *Chapter 4,* a review of successful RCT funding applications is undertaken. The decision-making behind intended design and analysis of these trials, and the driver of these, is identified. In *Chapter 5,* a survey of clinical trial statisticians is presented. The survey reports on the experiences, and management approaches, currently used within the UK Clinical Trials Units with regards to clustering effects and the learning curve in RCTs.

In *Chapter 6,* real example RCTs, each with different intervention types, are described in terms of how learning and clustering effects were managed during design and analysis. Selected examples are then used to illustrate the application of statistical methods, to explore the presence of learning and clustering, in the subsequent two chapters. In *Chapter 7,* the presence of clustering within an example RCT and the impact of clustering on trial conclusions is investigated. In *Chapter 8,* the presence of learning within an example RCT and its impact on trial conclusions is explored.

Finally, *Chapter 9* concludes this thesis with a summary of the main findings and ideas for future work.

Chapter 2 : Guidance on managing learning and clustering

2.1. Introduction

A number of guidance documents have been developed to support clinical trial design, conduct and reporting. The majority of these guidance documents target generic aspects of clinical trials, and as such are relevant to all trials. However, surgical trials have additional challenges, such as the learning curve and clustering, that should be considered. The extent that these are covered within existing guidance has not yet been reviewed.

The aim of this chapter is to identify and review existing guidance and consider their relevance to learning curves and clustering effects within surgical trials.

2.2. Methods

This work sought to include guidance documents developed to inform the design and analysis of randomised controlled trials (RCTs). Guidelines for inclusion in the cohort were identified by undertaking:

- An electronic search within the Equator Network (http://www.equator-network.org), an online library containing a comprehensive searchable database of reporting guidelines, using each of the search terms “surgery” and “statistic”. Documents that provided guidance specific to non-randomised studies, aspects of trial methodology or medical specialties that were not applicable, or focussed on applicable medical specialties, such as surgery, with no statistical scope were excluded.
- A targeted search of guidelines endorsed by leading UK funding bodies, regulators and medical journals such that they covered aspects of trial design, analysis and reporting.

Because of the nature of the search, full texts of identified guidelines were obtained to determine eligibility. Documents that provided guidance such that RCTs and statistical aspects were covered within their scope were included and reasons for exclusion were recorded.

Key criteria relevant to design and analysis of surgical trials, or trials of complex interventions, were identified a priori, see *Box 1*. Eligible documents were compared against these to identify gaps or inconsistencies in recommendations. Guidelines for reporting the aspects of design and analysis were also assessed against these criteria. Specific methods within the Guidelines related to analysing learning or clustering, at the centre or treatment provider level, were also collected. Documents were examined using NVivo qualitative data analysis software (QSR International Ply Ltd. Version 12, 2018).

Box 1: Key criteria to be considered within design and analysis

Design

- The appropriate trial design, such as an expertise-based design.
- Delivery of the intervention in terms of:
    - The health professionals delivering treatment.
    - The extent to which treatments are to be standardised.
    - The potential for change in delivery over time.
- Adjusting the sample size.
- Balancing treatment within centres and treatment providers.

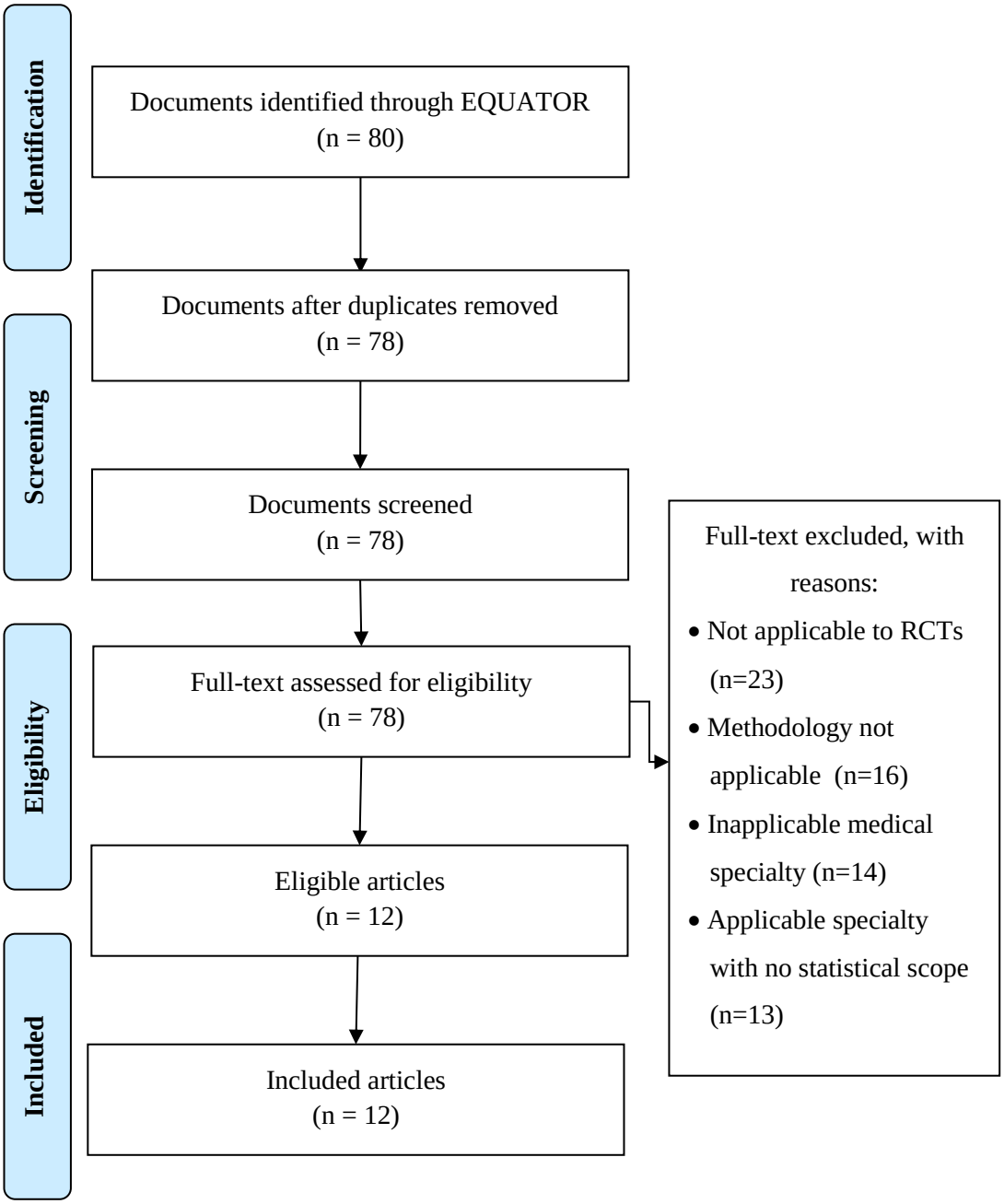
Analysis

- When the randomisation has been stratified.
- When analysing the primary outcome, such as adjustment.
- When there are multiple centres and/or treatment providers.

2.3.Results

2.3.1. Identifying eligible documents

Figure 1: Flowchart of eligibility for EQUATOR guidelines



The search within the EQUATOR website identified 80 documents: 36 (45%) were identified using the search term "statistic" and 44 (55%) using the search term "surgery". The search was conducted on 21st October 2021. *Figure 1* presents the flowchart of eligibility, with

reasons for exclusion where necessary. An additional 16 documents were manually identified from the targeted search (Funders: 2; Regulators: 6; Journals: 8, see *Appendix Material 1*). There were no duplicates between the two searches leaving a total of 28 eligible documents for review. *Appendix Material 1* provides the list of included documents.

2.3.2. Designing a trial with learning and clustering

Choosing a trial design

Eleven out of 28 documents (39%) provided guidance relating to trial design. See D1 in *Appendix Table 1*.

The options of trial design depend on the unit of randomisation and the intervention of interest. The key aspects of relevant designs are briefly summarised here. Many design options, and associated limitations, were discussed and no single document provided a single comprehensive summary.

In individually randomised trials, patients are the unit of randomisation. (13) When conducting these trials in surgery, differential expertise between the treatments being investigated can raise issues that can be alleviated by defining eligibility criteria for centres and surgeons, such as years in practice or number of interventions performed previously. (21, 22) However, applying criteria that are too strict may reduce the generalisability of trial results. (23) Instead, a statistical analysis of inter-rater reliability, between individual centres and surgeons, can provide understanding of any impact due to expertise. This type of analysis can be useful when considering rolling out the interventions into routine healthcare, see *Section 2.3.3*. (22)

In cluster randomised trials, groups of patients are the unit of randomisation. These designs are less common and are generally less efficient than individually randomised studies. They require more surgeons and introduce the potential for the treatment comparison to be

confounded by the delivery, despite inflating the sample size to account for the intraclass correlation coefficient (ICC). (13, 14, 24)

Expertise-based designs are a half-way house between individual and cluster randomised trials. Patients are individually randomised to surgeon, who each treat all patients with a single intervention. This can be the surgeon's preferred technique or are an intrinsic feature in trials comparing interventions delivered by different specialties. (14, 21) This design has the same limitations as cluster trials, and when a surgeon is only performing their preferred technique, shared waiting lists (14) and understanding how the treatment can be rolled out into routine healthcare can be a challenge. Resultantly, this design is relatively uncommon. (25, 26)

Tracker designs, proposed by *Ergina et al*, where new or evolving interventions can theoretically be developed within a single randomised study, and the incremental changes to the intervention tracked within the analysis, would be very challenging in practice. (14) Multiphase Optimization Strategy (MOST) designs, primarily developed for building and evaluating eHealth interventions, consist of a screening phase, a refining phase, and confirming phase and therefore may be particularly suitable for application to trials involving surgical interventions. (13)

Considering who will deliver the intervention

Thirteen out of 28 documents (46%) discussed the importance of deciding who will deliver the intervention. See D2 in *Appendix Table 1*.

Some variation in delivery, in part, will depend on the skill and training of those delivering the intervention. (14, 21, 27) As such, the selection of centres and treatment providers was a critical element of design discussed by a number of guidance documents. (16, 24, 28, 29) Any eligibility criteria for participating centres and treatment providers, and a description such as the degree to which they are typical, should be reported. (21, 23, 30)

Suggested selection criteria for centres related to caseload for the procedure under investigation and, likewise, ensuring sufficient numbers of the target population and facilities to deliver the trial. (21, 29)

No guidelines provided advice on selecting treatment providers. Treatment providers could be a limited group or all professionals offering the intervention. (31) If a limited group, guidance on selecting centres, and reporting requirements, may be looked upon as a proxy for triallists when deciding how to select treatment provider, for example caseload and ensuring specific qualifications. (21, 29, 30)

The results of the main trial should report on the number of centres and treatment providers performing each intervention. (30)

Ensuring that the intervention is standardised

Fifteen out of 28 documents (54%) discussed the importance of standardising the intervention. See D3 in *Appendix Table 1*.

Variation in delivery can be reduced by standardising all, or aspects of, the intervention of interest. Limiting variation in treatment delivery may be more desirable in an efficacy trial than a pragmatic, effectiveness study. (13, 15) In pragmatic trials. standardisation might consist of simply informing treatment providers to perform the treatment as usual. (21) Regardless of the stage, trial delivery should be similar at all centres (16) and designed such that a clear description of the procedures performed can be provided. (23, 32) Investigator meetings to prepare investigators and standardise performance were suggested by one document. (33)

Monitoring treatment adherence was an important aspect across documents. (15, 16, 21, 33, 34) Suggested methods included reviewing case report forms, videotapes and audiotapes,

extending to decertifying and excluding surgeons not submitting a videotape rated acceptable by an independent committee. (21)

Reporting in depth details of the intervention, and comparator, was required by a number of documents. Aspects required included technical procedures, full details on preoperative, intraoperative and postoperative care and the extent to which delivery were permitted to vary between participants, treatment providers and centres. (21, 23, 34)

Anticipating changes over time

Eight out of 28 documents (29%) discussed considering changes in delivery of the intervention over time. See D4 in *Appendix Table 1*.

Delivery may still vary irrespective of training, experience and other steps to enforce standardisation. The amount of variation will depend on the stage and technicality of intervention development. (13, 15, 21, 35) An important aspect of surgical evaluation across the guidelines was that delivery may change over time for pragmatic reasons, changes in external factors, or as a result of expertise developing during the study. (13-15)

Expertise can develop over a very long time and so requiring a set expertise level can slow the delivery of surgical trials. (14) Some guidelines discussed evaluating the learning curve within the trial, (15) and highlighted this was particularly important in earlier phase trials. (35) In trials comparing more established techniques, the statistical advantages and gain in 'internal validity' need to be considered against the loss of generalisability or 'external validity' of applying too much emphasis on the learning curve. (13)

Reporting learning curve assessment results was required by one document but this was limited to early phase studies. (35)

Estimating the sample size

Eight out of 28 documents (29%) discussed sample size. See D5 in *Appendix Table 1*.

A number of guidance documents highlighted the impact of failing to reduce variation within trial arms by standardising the intervention on the sample size and power calculation, where typical estimates assume that differences between the treatments across centres, or treatment provider, are unbiased estimates of the same quantity. (13, 16) In the presence of multilevel data structures, where variability in individual level outcomes can reflect higher level processes, calculations are more complicated. (16, 17, 27) To avoid associated imprecision in results, the sample size should adjust for any clustering effects as estimated by the intraclass correlation coefficient (ICC) and this should be reported in the main results paper. (21, 30) Conversely, two documents that discussed sample size did not comment on adjusting for clusters. (22, 29)

Ensuring balance of treatment within centre and treatment provider

Six out of 28 documents (21%) discussed ensuring that treatment allocations are equally distributed within centre. See D6 in *Appendix Table 1*.

Balancing treatment groups with respect to prognostic factors enhances trial credibility. (29, 36) Ensuring balancing of patients within centre was highlighted as important within many of the guidance documents, (16, 29, 36) and similar reasoning would lead surgical triallists to extend this to treatment provider which was not discussed within any document.

Balance can be achieved by stratifying the randomisation and stratifying by centre was a common topic, particularly when centre is expected to be confounded with other prognostic factors. (16, 29, 36) When there are too few patients per centre, stratifying by a larger unit, such as country or region, may be warranted. (36) Despite stratifying by treatment provider not being specifically addressed within the documents, in some circumstances it may be

desirable to stratify for more than just both centre and treatment provider, or treatment provider alone, where numbers allow. (36) The use of more than two stratification factors is rarely necessary. (16)

2.3.3. Analysing a trial with learning and clustering

When the randomisation was stratified

Two out of 28 documents (18%) provided guidance on adjusting the analysis following stratification. See A1 in *Appendix Table 1*.

Stratifying randomisation and subsequently adjusting the analysis, are complementary methods of accounting for prognostic factors, unless the stratification factor was chosen for administrative reasons only. (16, 36)

Two documents discussed the issue of adjusting for too many, or too small, strata in the analysis, for which there is no best solution. (16, 36) When included in the randomisation scheme, ignoring centres or adjusting for a large number of small centres might lead to unreliable estimates of the treatment effect and p-values. (36) At best, using an unadjusted analysis should be supported by sensitivity analyses that indicate trial conclusions are not affected because of this. (36) As in previous sections, the statistical justifications for including centre could be considered to also include treatment provider in surgical trials but no guidance required this specifically.

When analysing the primary outcome

Two out of 28 documents (18%) provided guidance on adjusting the primary outcome analysis. See A2 in *Appendix Table 1*.

Unexplained differences between treatments, for example between adjusted and unadjusted analyses, can jeopardise the trial results. (36) For this reason, when the primary outcome is expected to be influenced by centre or treatment provider, an adjustment should be planned. When the potential value of an adjustment is in doubt, such as little existing prior knowledge, the primary analysis should be unadjusted analysis, supported by an adjusted analysis. (16, 36) In general, larger datasets generally support more factors than smaller ones and results based on simpler models are generally numerically stable, the assumptions underpinning the statistical model are easier to validate and more generalisable. (36)

Analysing multi-centre trials

Six out of 28 documents (21%) provided guidance on analysing multi-centre trials. See A3 in *Appendix Table 1*.

Investigations into heterogeneity of the main treatment effect across centre and/or treatment provider were covered by a number of documents. (15, 16, 21, 34, 35) Further, the main trial publication should report methods to adjust for, and results into, clustering by centre or treatment provider. (21, 30) These investigations are critical when a positive treatment effect is found and there are appreciable numbers of subjects per centre. (16) In the simplest multi-centre trial, a single investigator recruits and is responsible for all patients within a single hospital, such that centre is identified uniquely by hospital. When the definition of centre is ambiguous, such as a single investigator recruits from several hospitals or a clinical team recruits from numerous clinics, the protocol should provide a definition. (16, 34)

Quantitative approaches may comprise graphical display of the results of individual centres, such as forest plots, or analytical methods, such as a significance test although this generally has low power. (16) One document stated that investigations use a model which allows for centre differences but no interaction terms. (16) Fixed or mixed effects models can be used,

although mixed models are especially relevant when there is a large number of centres. (16, 34)

Methods for investigating the learning curve

Four out of 28 documents (14%) provided guidance on analysing the learning curve within centre and/or treatment provider. See A4 in *Appendix Table 1*.

Reporting of continuous quality control measures can be useful for all phases of trial, particularly early phase surgical trials. (15, 35) Time series and longitudinal models or multilevel models can be used to analyse long and short sequences of data respectively. (13, 27) Simpler exploratory methods such as cusum plots enable centres or surgeons to be compared against themselves, as opposed to each other, which can be preferable to surgeons. (15, 35)

Methods for investigating clustering

Five out of 28 documents (18%) provided guidance on investigating clustering due to centre and/or treatment provider. See A5 in *Appendix Table 1*.

Hierarchically structured data, such as patients within surgeon, can be analysed using multilevel models or generalised estimating equations (GEE). (13, 30) Multilevel models are subject-specific models whereas GEEs are population average. For multilevel models: fixed, random or mixed effects can be specified to account for clustering (30) and different types of these models allow for flexible data structures. (27)

For ordinary linear models, the treatment effect estimate is likely to be similar but not necessarily identical for adjusted and unadjusted models, see page 13. Adjusted analyses are more efficient, and so a less significant result for unadjusted should not be a concern. For

generalised linear or non-linear models, adjusted and unadjusted treatment effects may not have the same interpretation and may provide different results. (36)

2.4. Discussion

Triallists should consider the impact of learning and clustering when designing and analysing randomised surgical trials. Considerations should be incorporated into reporting to aid interpretation and applicability of trial results. This chapter provides the first review as to the extent that existing guidance covers these important considerations. Existing guidance documents are identified and summarised, with a focus on aspects relating to learning curves and clustering effects and their application to surgical trials. Not all documents were written specifically for surgery, yet all contain aspects that can be applied to surgery, for example the role of centre in delivery of treatments in drug trials is not dissimilar to the role of surgeon in delivering a surgical trial. Twice the number of identified documents targeted design aspects than analysis. Whilst a good analysis cannot rescue a poor design, and this may have led to a larger focus on design on guidance for triallists, there is a notable dearth of analysis guidance available that should be addressed. In addition, there is also scope for guidance on study conduct.

A number of guidance documents acknowledged the importance of the surgical learning curve, or delivery changing over time, within design and analysis, particularly in early phase surgical trials or when the interventions differ in their technicality (13-15, 21, 35) Yet there was little coverage within reporting standards to reflect this, with the surgical learning curve analysis stated as only necessary in early phase, and not necessarily randomised, trials (35) and broader RCT reporting guidelines only requiring differential expertise be addressed in the discussion. (21) Lack of clear standards may lead to reporting how delivery of intervention changes over time, despite its importance, being generally under-recognised in the literature.

Clustering, at the centre level, was well covered within the design, analysis and reporting guidance. However, the were inconsistencies with regards to the treatment provider coverage. For example, reporting required eligibility of treatment provider be covered, yet no guidance on the design or analysis covered this. (21, 30) However, this may be due to the original guidance largely not being written specifically for surgery, or indeed complex interventions, where these effects may be more prominent. The role of centre within conventional drug trials could be extended to provide guidance on the role of treatment provider in surgery trials. (13, 16, 29) A more focussed guidance document that covers the design and analysis of randomised surgical trials, or intervention trials, could address this discrepancy to improve quality of understanding and awareness of these issues.

When reviewing this chapter, it is important to consider the limitations of this summary. First, country specific guidance beyond the UK, such as United States Food and Drug Administration, were not included. However, international documents that are applicable to other countries, including the UK, were obtained, such as ICH which are followed globally and EMA which are adopted within Europe. Second, only four guidance documents, developed by the same research group, were written specifically for surgery and not written specifically for RCTs, which may explain the lack of specific coverage of the surgeon in the wider set. (14, 15, 35, 37) Third, very little of the guidance documents covered statistical aspects, leaving a triallist to extend the centre-drug connection to surgeon-intervention themselves. (16, 36) A statistical guidance document that covers randomised surgical trials in more depth would help triallists, in particular statisticians, and the IDEAL framework provides a good basis for this development to be integrated. (14, 15, 35, 37)

2.5. Conclusions

This review has identified that guidance on trial design is better covered within existing guidance than trial conduct and analysis. However, no single and complete guidance document

exists that cover aspects of learning and clustering leaving triallists to have to access multiple documents to gain full understanding of these considerations. This has the potential to lead to inconsistencies in practice, which is further explored in *Chapter 3, Chapter 4* and *Chapter 5.* The impact of failing to consider clustering and learning in analysis is investigated in *Chapter 7* and *Chapter 8* respectively.

Chapter 3 : Reporting considerations for learning and clustering

3.1. Introduction

In order to investigate approaches for handling learning curves and clustering effects in randomised surgical trials, a review of the published literature was undertaken. This review aimed to investigate and establish reporting standards and practice for the consideration in these effects. This review was thus designed to ensure that it was wide to ensure appropriate coverage of the literature.

This chapter begins by providing the justification for this review of the published literature (*Section 3.1*). In *Section 3.2*, the methods used to perform this review are defined. Results are provided in *Section 3.3*.

The work in this chapter has been published in the Journal of Clinical Epidemiology (JCE) and I am first author. (25)

3.1.1. Importance of complete reporting

It is well recognised that clinical trials often fail to report important features of design and analysis that are necessary to ascertain the methodological quality of the trial. In particular surgery. (21, 30)

Communicating these considerations through complete and transparent reporting can aid appraisal and interpretation by the wider surgical community. Reporting standards for trials of non-pharmacologic treatments, such as surgery, have been established. Amongst their requirements, reporting of items specifically relating to learning curves and clustering are recommended. (21, 30)

3.1.2. Aims

Learning curves and clustering have so far been investigated in isolation, often within specific fields and including studies of observational design. (19, 20, 38) The objective of this work was to provide an up to date and comprehensive overview of current practice when reporting randomised surgical trials, regarding the management of surgical learning and clustering effects in design and analysis.

3.2. Methods

3.2.1. Terminology

In this review, the term "treatment provider" and "surgeon" are used interchangeably to mean the individual administering the treatment (or intervention) to a consenting patient.

The term "centre" is used to indicate the designated medical facility or site used to conduct the clinical research e.g. hospital or GP practice.

The term "surgical intervention" is defined as an intervention that involves physically changing body tissues and organs through manual operation, such as cutting, abrading, suturing and the use of lasers. (12)

3.2.2. Included studies

This review sought to review reports of randomised surgical trials within the wider surgical literature. Manuscripts for inclusion in the cohort were identified by undertaking an electronic search using Scopus from a subset of journals. These journals were identified as the ten leading English-language general surgical journals (1-10, *Box 2*) (39) plus six general medical journals (11-16, *Box 2*). The rationale for selecting leading surgical and general medical journals was based on the assumption that they endorse high standards of RCT reporting. The search was deliberately designed to be wide to ensure appropriate coverage of the literature as statistical

aspects of surgical trials tend to be underreported. (40) An informal, feasibility search for eligible trials within these journals identified 124 trials published in 2014 within the surgical journals, with Annals of Surgery contributing half. For general medical journals, 23 trials were published in 2013 and 2014. A further 18 surgical RCTs were identified within the entire HTA monograph series to January 2015.

Box 2: List of journals searched for main trial publications

1	Annals of Surgery
2	British Journal of Surgery
3	Archives of Surgery - now JAMA Surgery
4	Surgery (United States)
5	American Journal of Surgery
6	Journal of American College of Surgeons
7	Current Problems in Surgery
8	American Surgeon
9	Australia and New Zealand Journal of Surgery
10	Surgery Today
11	British Medical Journal (BMJ) [A]
12	Journal of the American Medical Association (JAMA) [A]
13	The Lancet [A]
14	The New England Journal of Medicine
15	Trials
16	Health Technology Assessment

[A] Including specialty titles

Manuscripts of RCTs evaluating a surgical intervention, or a non-surgical intervention requiring surgery to be administered, published from 1st January 2014 to the date the search was conducted (11th February 2016), were eligible. Duplicate publications, secondary

analyses, and interim reports of RCTs were excluded. All eligible RCTs were included in the cohort.

3.2.3. Data extraction

Selected journals were screened for RCTs that meet the eligibility criteria. *Appendix Material 2* provides the search strategy for Scopus. Manuscripts were screened to identify those eligible for selection. Due to the nature of the intervention of interest, full texts were screened to determine eligibility. When suitability was unclear a second reviewer (CG) was consulted.

A data extraction form was developed by two project team members (EJC and CG), revised based upon feedback from GB, JAC, and JMB, and subsequently piloted on thirty articles prior to roll out to all articles, see *Appendix Material 3.* Data were extracted from all articles by a single assessor (EJC). Data extracted were quality checked through double data extraction by a second reviewer (ARH) on 10% of the articles. An error rate was specified a priori such that if greater than 5% across all fields then a further 10% would be checked until the error rate was below 5%. Data were extracted from all published materials (main trial report and, where applicable, supplementary material).

Data were extracted on generic trial design e.g. randomisation details and statistical analyses related to clustering and learning at a centre and surgeon level. Pre-determined centre and/or surgeon credentials and variables relating to surgical learning or clustering, either as a definitive outcome or as a variable of interest, such as duration of operation or the number of operations by surgical level were collected.

3.2.4. Statistical analysis

Quantitative items were summarised using descriptive statistics; no formal statistical comparisons were undertaken. SAS 9.3; SAS Institute Inc., Cary, NC, USA was used. Open

textual; responses were categorised using NVivo qualitative data analysis software (QSR International Pty Ltd. Version 10, 2012).

3.3. Results

3.3.1. Article details

The search identified 874 reports (398 in 2014; 446 in 2015; and 30 in 2016 – to date of extraction 11th February 2016) of which 247 were eligible. *Figure 2* provides the PRISMA flow diagram. *Appendix Material 4* and *Table 1* provide a list of eligible studies and summarise the cohort demographics respectively. Two surgical specialty journals and three general medical journals contributed the majority of eligible papers (*Table 1***)**. When reported (167/247, 68%), over half of eligible trials were European funded (n=92/167, 54%), and over a quarter were North American funded (n=48/167, 29%).

Twenty-five articles were randomly selected from the eligible manuscripts for double data extraction. Of 1025 variables checked, 12 errors were identified (1.2%).

Figure 2: Flowchart of eligibility for main trial publications

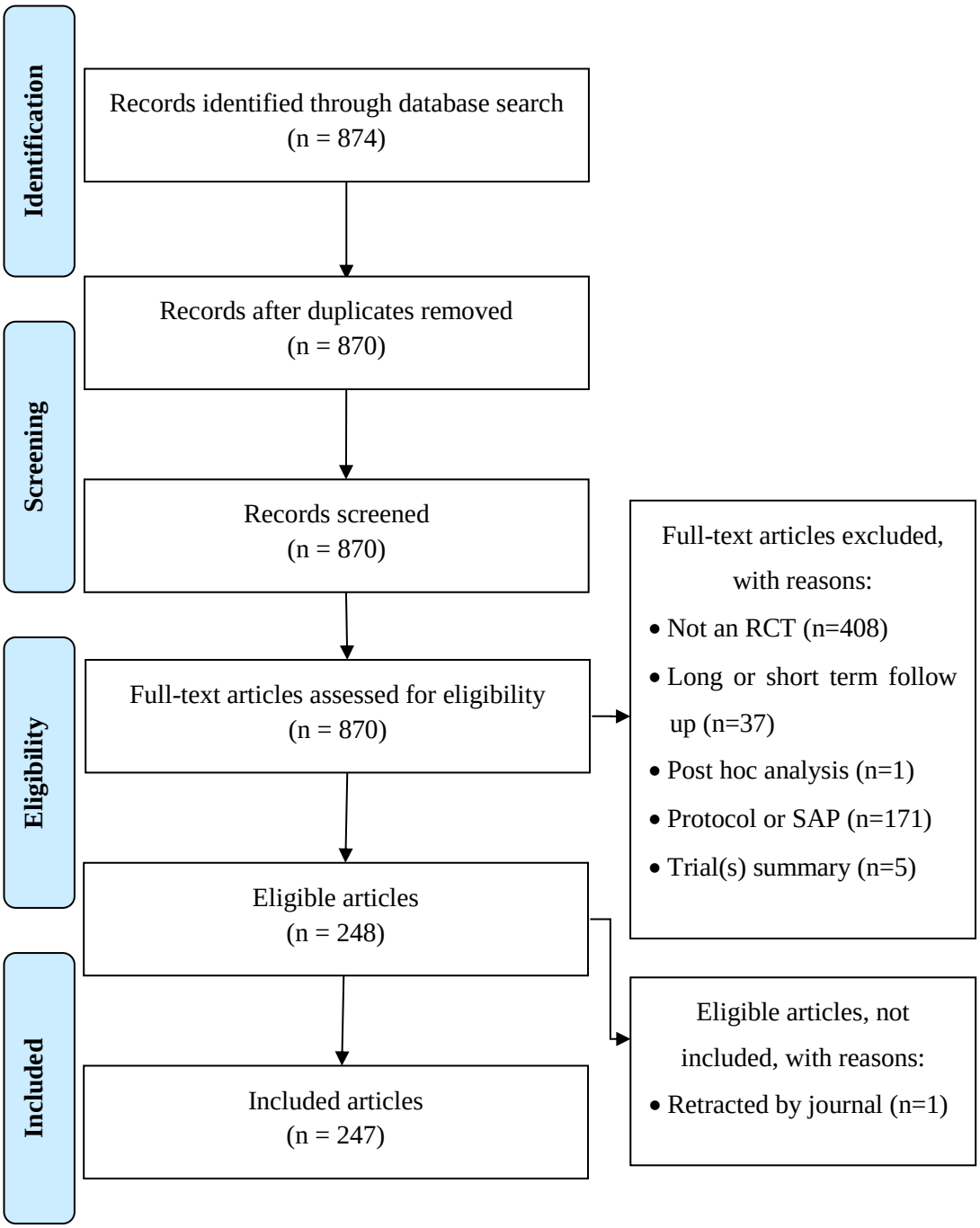


Table 1: Demographics of included studies

Item	Category	n	N	n/N%
Year published	2014	101	247	41%
	2015	140	247	57%
	2016	6	247	2%
Journal	Annals of Surgery	34	247	14%
	The New England Journal of Medicine	28	247	11%
	British Journal of Surgery	21	247	9%
	Journal of the American Medical Association	20	247	8%
	The Lancet	19	247	8%
	Journal of the American College of Surgeons	16	247	6%
	Surgery (United States)	15	247	6%
	ANZ Journal of Surgery	12	247	5%
	Surgery Today	11	247	4%
	Health Technology Assessment	10	247	4%
	The Lancet Oncology	9	247	4%
	JAMA Surgery	8	247	3%
	American Journal of Surgery	6	247	2%
	JAMA Opthalmology	6	247	2%
	Trials	6	247	2%
	American Surgeon	5	247	2%
	JAMA Facial Plastic Surgery	5	247	2%
	BMJ (Online)	3	247	1%
	The Lancet Neurology	3	247	1%
	BMJ Open	2	247	1%
	JAMA Dermatology	2	247	1%
	The Lancet Diabetes and Endcrinology	2	247	1%

Item	Category	n	N	n/N%
	BMJ Supportive and Palliative Care	1	247	<1%
	JAMA Otalaryngology (Head and Neck)	1	247	<1%
	The Lancet Respiratory Medicine	1	247	<1%
	Surgery (United States)	1	247	<1%
Supplements [A]	Appendix	99	237	42%
	Protocol	50	237	21%
Funder origin	Asia	19	167	11%
	China	7	167	4%
	Japan	6	167	4%
	Malaysia	1	167	1%
	South Korea	5	167	3%
	Australia	10	167	6%
	Australia	9	167	5%
	New Zealand	1	167	1%
	Europe	90	167	54%
	Belgium	1	167	1%
	Denmark	3	167	2%
	Finland	3	167	2%
	France	5	167	3%
	Germany	7	167	4%
	Ireland	2	167	1%
	Italy	2	167	1%
	Netherlands	14	167	8%
	Norway	3	167	2%
	Spain	1	167	1%
	Sweden	6	167	4%

Item	Category	n	N	n/N%
	Switzerland	2	167	1%
	United Kingdom	40	167	24%
	Multiple countries	1	167	1%
	North America	48	167	29%
	Canada	10	167	6%
	USA	38	167	23%
	None	10	.	
	Not clear	70	.	

[A] Not applicable to Health Technology Assessment series as publications consist of a single document containing all appendices.

3.3.2. Trial rationale and design

Design features and characteristics of the trials are summarised in *Table 2*. Included trials were typically of a parallel (n=231, 94%), two-armed (n=224, 91%) design. 64% described approaches to blinding trial personnel (n=157/247).

Table 2: Trial design features and characteristics

Item	Category	n	N	n/N%
Trial type	Definitive study	240	247	97%
	External pilot or feasibility (41)	7	247	3%
Type	Cluster	3	247	1%
	Crossover [A]	2	247	1%
	Factorial	6	247	2%
	Parallel	231	247	94%
	Sequential	5	247	2%
Number of arms	2	224	247	91%
	3	16	247	6%
	4	6	247	2%
	6	1	247	<1%
Some trial personnel blinded	Yes	157	247	64%
	No	47	247	19%
	Not reported	43	247	17%
Expertise Design	Pure – professionals delivering only one intervention [B]	4	247	2%
	Hybrid – some professionals could deliver both [C]	1	247	<1%
Intervention of interest	Surgery occurred but was not intervention of interest	105	247	43%
	Surgery occurred and was the intervention of interest	142	247	57%
Comparator when surgery was intervention of interest? [D]	Surgery	111	142	78%
	Medical	10	142	7%
	Other e.g. active monitoring	25	142	18%

Surgical comparison in trials comparing two surgeries?	Comparing different components of the same intervention	68	111	61%
	Different surgical interventions	38	111	34%
	Different time points of the same intervention	5	111	5%

[A] Includes designs in which each participant receives both interventions

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

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

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

Within the cohort, over half of the trials were reported as multi-centre (n=130, 53%) and two-thirds multi-surgeon (n=162, 66%, *Table 3*).

Table 3: Randomisation considerations

Item	Category	n	N	n/N%
Multiple or single centre trial	Multiple	130	247	53%
	Single	101	247	41%
	Not reported	16	247	6%
Multiple or single treatment provider trial	Multiple	162	247	66%
	Single	22	247	9%
	Not reported	63	247	25%
Randomisation stratified	Yes	123	247	50%
	No	124	247	50%
If yes, randomisation stratified by	Centre and treatment provider	2	123	2%
	Centre	77	123	63%
	Treatment provider	8	123	6%
	Neither	36	123	29%
Allocation of treatment provider	*Pure – professionals delivering only one intervention*			
	Defined by research question	1	4	25%
	Preference	1	4	25%
	Randomised	2	4	50%
	Hybrid – some professionals could deliver both			
	Preference	1	1	100%

Very few trials utilised an expertise-based design, (42) where the health professionals deliver only one of the comparators (n=5, 2%). One of these used a hybrid design where some health professionals could deliver both interventions. Treatment providers were allocated to arm

based on: preference (n=2, of which 1 was a hybrid); randomisation (n=2) and the research question (n=1) (*Table 2*).

3.3.3. Intervention of interest

Surgery occurred and was the intervention under evaluation in 57% of trials (n=142/247, *Table 2*). Three quarters (n=111, 78%) of these also had a surgical comparator, the majority of which compared different components of the same intervention (n=68/111, 61%), one-third compared different surgical interventions (n=38/111, 34%) and a small number compared different time points of the same intervention, such as early or delayed surgery (n=5/111, 5%). *Box 3* gives an example from each category where surgery was a comparator. 43% (n=105/247, *Table 2*) were trials in which surgery was not the intervention under evaluation but instead delivered as a co-intervention, for example a trial of neoadjuvant chemotherapy and surgery where surgery was the same in both arms.

Box 3: Examples of surgical comparisons by category

Different components of the same intervention:

Patients underwent transcatheter aortic valve replacement with (A) balloon expandable device or (B) self-expandable device *[Abdel-Waheb 2014]*

Different surgical interventions:

Patients received (A) perineal dissection or (B) abdominal dissection *[Abo-Ryia 2015]*

Different time points of the same intervention:

Patients received lumbo peritoneal shunt surgery (A) within one month or (B) delayed for three months *[Kazui 2015]*

3.3.4. Centre and surgeon credentials

Pre-defined centre and surgeon credentials were reported in 41% of trials (n=101/247, *Table 4*). This included 96 (of 162) multi-surgeon trials, with common definitions being a set prior number of cases (n=27) or a specific level or job role e.g. consultant (n=22). Fourteen trials reported criteria at centre level, of which nine reported these alongside surgeon criteria. Examples of reported criteria are summarised in *Table 4*.

Table 4: Centre and surgeon credentials

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

3.3.5. Randomisation

Table 3 provides a summary of randomisation considerations. Half stratified the randomisation (n=123, 50%), using methods such as block randomisation or minimisation. Seventy-nine of the 130 multi-centre trials (61%) and 10 of the 162 multiple surgeon trials (n=10/162, 6%) stratified accordingly. Of the surgeon stratified trials (n=10), half were trials comparing different components of the same intervention (n=5) (*Table 5*). Two multi-surgeon trials stratified by both centre and treatment provider and almost half stratified by neither (n=75/162, 46%). *Table 6* summarises the stratification approach within multi-surgeon trials.

Table 5: Adjustment and stratification by comparator of interest in multi-surgeon trials

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

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

Table 6: Multiple surgeon trials stratification approach for centre or surgeon

		Stratification factor							
		Centre and surgeon		Centre only		Surgeon only		Neither	
Centres	**N**	**n**	**n/N%**	**N**	**n/N%**	**n**	**n/N%**	**N**	**n/N%**
Multiple	130	2	2%	77	59%	5	4%	46	35%
Single	28	.	.	.	.	2	7%	26	93%
Not reported	4	.	.	.	.	1	25%	3	75%

3.3.6. Considerations of learning and clustering of centres and surgeons

Variables reported relating to learning (*Box 4*) were: background or level of surgeon or centre (n=14, 5%, of which one gave both); experience in years (n=5, 2%); number of operations by surgeon level (n=13, 5%); or over time (n=1, <1%). Operation time was most commonly reported (n=82, 33%). Variables relating to clustering (*Box 4*) were: number of patients by region (n=1, <1%); centre (n=39, 16%); surgeon (n=13, 5%); the number of surgeons per centre (n=1, <1%) and the overlap of surgeons between arms (n=2, 1%).

Box 4: Descriptive items

Centre:

Number of patients by centre (n=39, 16%); background or level of centre (n=2, 1%); number of surgeons by centre (n=1, <1%).

Surgeon:

Operative time (n=82, 33%); background or level of surgeon (n=13, 5%); number of operations by surgeon level (13, 5%); surgeon caseload (n=13, 5%); experience in years (n=5, 2%); overlap of surgeons between arms (n=2, 1%); number of patients over time (n=1, <1%).

Other:

Number of patients by region (n=1, <1%).

Of the 79 multi-centre trials that stratified by centre, one third (n=25) reported within centre descriptive data, for example caseload. Likewise, of the 10 stratified multi-treatment provider trials, half (n=5) reported descriptive data (*Table 7*).

Table 7: Stratification factor by descriptive items of centre and treatment provider

		Reported descriptive items e.g. case volume of surgeon			
		Yes		**No**	
Stratification factor	**N**	**n**	**%**	**n**	**%**
Centre	79	25	32%	54	68%
Treatment provider	10	5	50%	5	50%

Outcomes potentially relevant to clustering or learning, for example length of operation, are presented in *Box 5*. 80% (198/247) reported on at least one outcome relevant to clustering or

learning curves with the most commonly reported being safety events (n=129, 51%) and infection (n=46, 19%).

Box 5: Outcomes summary

283 outcomes reported within 198 trials. Types of outcomes are as followed:

Infection (n=46, 19%)

Surgical site infection (n=22, 9%); wound infection (n=15, 6%); other (n=9, 4%).

Recovery (n=70, 28%)

Length of stay (n=52, 21%); length of stay in intensive care unit (n=18, 7%).

Operative time (n=23, 9%)

Safety including complications, adverse events and serious adverse events (n=129, 51%)

Surgeon outcomes such as technical performance and perceived difficulty (n=13, 5%)

Centre outcomes such as recruitment totals (n=2, 1%)

3.3.7. Analysis adjustment of centres and surgeons

Centre or treatment provider when used to stratify the randomisation process, were also used to adjust the analysis (*Table 8*) in one third of trials (n=26/79, 33%). Of the ten trials that stratified by treatment provider, four made analysis adjustments. One third of multi-centre trials (n=45/130, 35%) and almost 90% of multi-treatment provider trials (n=140/162, 86%) neither stratified randomisation nor made analysis adjustments.

Table 8: Stratification of randomisation by analysis adjustment by centre and treatment provider

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

39 considered centre effect through analysis of primary or secondary outcomes (N=247, 16%, *Table 9*). When reported, adjustment using a random effect was more common (n=16) than fixed effects (n=1). Adjustments were applied to all outcomes in one third of trials, and to primary outcome only in almost half. Other approaches included: a sensitivity analysis excluding the centre with the largest number of participants; and using centre as a predictor to impute missing value, see *Table 8*.

20 considered surgeon effect through analysis of primary or secondary outcomes (N=247, 8%, *Table 9*), with the majority of these were trials comparing different surgical interventions (n=8, 40%) (*Table 5***)**. Adjustments were applied to all outcomes in one third of trials, and to primary outcome only in 40%. Other approaches were to explore safety of surgeons in delivering interventions in a separate paper and to consider "run in" patients where the first 100 patients were randomised separately in analysis, see *Table 8*.

Table 9: Statistical adjustment for multiple centre and surgeon effects in primary or secondary analyses

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

3.4. Discussion

This review examines methods for addressing learning and clustering effects within a large cohort of 247 randomised surgical trials. Most commonly, learning effects were addressed in the design of the trial by surgeon or centre participation requirements e.g. number of previous operations. Expertise based studies were rare, although some may have been expertise-based in delivery but not reported as such. (12) One study conducted a formal investigation of the learning curve by using a time varying treatment effect. Clustering was also most commonly accounted for in the design stage, by stratifying the randomisation process, by centre and/or surgeon. However, in most cases the analysis was not then adjusted to reflect this. (43)

Numerous examples in the literature demonstrate the presence of a learning curve and investigate the impact on outcomes over time. (44, 45) In the surgical field, the appropriateness of making considerations for surgeon in an individual trial should be considered against how commonplace and stabilised the procedure or intervention are within routine practice. For example, consideration may be given to whether the trial is comparing established practices, established practices with minor differences or entirely different or radical new procedures. Formal analysis of surgical learning was rare, when triallists consider the learning curve to be of interest, for example early phase studies involving radical new procedures, established statistical methods that allow the learning profile to be explored may be considered. (46)

Approaches to manage clustering at the surgeon level was less prevalent than at centre level within this cohort. This may be appropriate reflecting on the nature of the interventions being compared and their routine use. Impact of care bundles, for example pre and post-surgical care, may be considered to exert a greater influence on outcomes than individual surgeon. These aspects of care are typically centre driven effects. A large cohort analysis of cardiac patients determined that 95% of variation in the outcome of interest was explained by patient risk factors. Surgeon and centre contributed only 2-3% respectively. (47) This raises the question of the importance of adjusting for surgeon particularly where the volume of data available limits the extent of modelling techniques. It is important to note that when surgeon and/or centre are prognostic indicators in a trial, the randomisation of the trial will often be balanced for this, commonly through stratifying the randomisation process. (16, 28, 36) However, the subsequent analysis should be adjusted for these chosen stratification factors. Failure to adjust following stratification can inflate p-values and confidence interval widths potentially creating erroneous conclusions of no treatment benefit. (43) Within this cohort, one third of centre stratified trials and 40% of surgeon stratified trials reported making necessary adjustments to the analysis.

This review has identified potential deficiencies in the design and analysis of surgical trials. The regulatory governance of surgical trials is not comparable to pharmaceutical trials, however many of the requirements are directly relevant. (16, 28, 34, 36) The *ICH E9: Statistical principles for clinical trials* document discusses reasons for conducting multi-centre trials and the importance of defining the centre appropriately either by centre or investigator. (16) This is directly applicable to surgical trials. Further guidance states that the potential for differential treatment effects across centres should be explored, with individual centre results being reported and treatment-by-centre interaction considered in the absence of homogeneity. (16, 34) Our results show that practice does not match this guideline, with one third of multi-centre studies, and 13% of multi-surgeon studies, reporting approaches to check for differential outcome effects, or justifying not doing so. It is important to remember that heterogeneity may be caused by factors not related to the surgeon. Heterogeneity may be explained at the centre level for example by differences in patient demographics, or at the level of the treatment provider for example due to variation in case mix complexity. Existence of heterogeneity between centres has implications for generalisability of study results and should be routinely investigated to appropriately consider generalisability. There is an absence of guidelines focusing on learning curves. This may be due to guidance stemming from medicinal trials or expectations that learning curves are either not expected to be apparent if the trial is comparing commonly used practices or suitably addressed in training prior to trial commencement and selection of treatment providers. Alternatively, it may be because it is difficult to measure surgeon expertise, methods are often imperfect and subject to other influences, such as case mix.

The need for transparency around learning curves and clustering are highlighted within the guidelines on reporting of non-pharmacological interventions. (21, 30) This review identified poor adherence to these reporting guidelines with key requirements missing or only partially reported. Coupled with the poor adherence to good statistical practice guidelines, (16) limitations in reporting may strengthen the concerns by health professionals that surgical

research is of a poor quality, as this can ultimately lead to ill-founded clinical decisions. (21, 30)

When interpreting these results, it is important to consider the limitations of this review. Firstly, this cohort was restricted to top surgical and medical journals, while advantageous as it provides a wide variety of trials by surgical discipline and by geographic location, these trials are more likely to be of a higher quality and demonstrate better methodological practice due to wider adoption of reporting guidelines. (48) Secondly, this cohort is cross-sectional and therefore does not consider changes over time. However, due to the recent growth in surgical trials, and the establishment of reporting guidelines for trials such as surgery, it is likely therefore that little is to be gained from reviewing more dated literature. Indeed, due to recent advances in the development of guidelines, as described in *Chapter 2*, conducting this review on a more recent sample may indicate different results to those presented here. Finally, drawing conclusions based on published manuscripts may be hindered by a lack of transparent reporting. Due to word count constraints and within journal requirements, authors may not have been able to fully report methods used although all supplementary material was included where available. Further insight into methods used could have been obtained by interviews with authors although this would be resource intensive. Further insight into current practice could be informed by contact with current surgical triallists and statisticians, or by exploring trial documentation that may not be published, such as grant applications or protocols. This is further explored in *Chapter 4* and *Chapter 5*.

3.5. Recommendations

Fundamental to the design and analysis of a trial is the understanding of the trial objectives. Many multi-centre trials have multiple centres not due to interests in how treatment effects vary by centre or surgeon. Instead, due to logistical considerations to provide a better basis for the subsequent generalisation of its findings and ensure sufficient availability of the patient population. Considerations and recommendations for design and analysis are presented for

surgical learning and clustering in *Box 6* to *Box 9* based on current guidelines and recommendations through example scenarios. (16, 21, 28, 30, 34, 36, 43) These aspects of trial design and analysis demonstrate the need for, early and continued expert statistical input. (19)

3.6. Conclusions

Considerations for surgical learning and clustering effects in published manuscripts is often unclear. Methods are varied and demonstrate poor adherence to established reporting guidelines. It is recommended that researchers consider these issues on a trial-by-trial basis, and report methods or justify where not needed to inform interpretation of results. Early, and continued, statistical input is essential to support the applications of appropriate methods when necessary.

Box 6: Scenarios for considering learning curve effects at design and analysis

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

Box 7: Recommendations for learning curve consideration

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

Box 8: Scenarios for considering clustering effects at design and analysis

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

Box 9: Recommendations for clustering consideration

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

Chapter 4 : Considering learning and clustering during trial design

4.1. Introduction

In *Chapter 3*, a lack of adherence to reporting guidelines was established (21, 30) which led to difficulties in concluding practice based on published manuscripts. Investigating beyond published manuscripts was recommended. Grant applications of randomised surgical trials were identified as a source of unpublished and detailed information to bridge the gap in understanding practice in accordance with what is published and what is being done. This review aimed to complete this gap in knowledge.

This chapter begins by providing further justification for this review of grant applications (*Section 4.1*). In *Section 4.2*, the methods used to perform this review are defined. Results are provided in *Section 4.3*.

The work in this chapter has been published in the Journal of Clinical Epidemiology (JCE) and I am first author.

4.1.1. Designing trials to account for learning and clustering

Recognition and management of learning curves and clustering within clinical trials is recommended, (16) and may have increased relevance within the surgical field, dependent upon the interventions being investigated and their routine use. (12, 14, 16-18)

It is important therefore to consider the significance of these aspects at trial outset, to ensure that the resulting trial is conducted and analysed with the highest possible rigour. However, main trial publications often do not report considerations and justifications for selected

approaches, see *Chapter 3*. To overcome this limitation, a cohort of applications for randomised surgical trials funded by the NIHR are investigated.

4.1.2. Aims

This review will determine how learning and clustering by centre and surgeon are managed at the design stage and accounted for in the intended analysis. It will also provide insight into who drives the decision-making for these: the funder, guided by reviewers and panel members, or the researcher. The aim is to provide a more detailed insight into current practice with regards to planning for, and acknowledging, the presence of learning and clustering at the design stage.

4.2. Methods

4.2.1. Included studies

Trials that had received funding from the NIHR from two funding streams, the Health Technology Assessment (HTA) programme (49) and Efficacy and Mechanism Evaluation (EME) programme (50) in the UK, from 2012 to 2016 were eligible. Research projects funded by these programmes are either in response to a commissioning brief or an open investigator led call. These funding streams were chosen as they are known to endorse high quality research and were actively funding surgical research during this time. (6) An initial unpublished search indicated that this period would provide a reasonable cohort size to establish current practice. All randomised trials where the patient pathway involved a surgical intervention of any kind were eligible for inclusion.

4.2.2. Documents for review

The NIHR HTA and EME funding process involves a two stage, peer reviewed application process. Protocols and the commissioning brief (where applicable) were obtained from the open access NIHR Journals Library. (51) The NIHR Evaluation, Trials and Studies

Coordinating Centre (NETSCC) provided documentation not publicly available: project descriptions and applicant responses to reviewer comments.

4.2.3. Data extraction

The data extraction form previously developed, see *Chapter 3*, was adapted for use on this cohort by EJC and CG and approved by GB, JAC, and JMB, see *Appendix Material 5*. The extraction form was piloted on five applications initially and, as no further amendments were required, subsequently used on all applications by a single assessor (EJC). Data extracted were quality checked through double data extraction by a second reviewer (ARH) on 10% of all applications. A discrepancy rate was specified a priori such that if greater than 5% across all fields then a further 10% would be checked until the rate was below 5%. Discrepancies were jointly reviewed and agreement reached, if agreement was not achieved then a third reviewer (CG) was consulted.

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

4.2.4. Statistical analysis

Quantitative items were summarised using descriptive statistics; no formal statistical comparisons were undertaken. Data was analysed using SAS 9.3; SAS Institute Inc., Cary, NC, USA. Open textual data items; were categorised using NVivo qualitative data analysis software (QSR International Pty Ltd. Version 10, 2012). A confidentiality agreement with the NIHR Evaluation, Trials and Studies Coordinating Centre was signed prior to receiving the documentation. The raw data cannot therefore be made publicly available and text extracts have been anonymised by removal of treatment or condition identifiers. Deleted text is

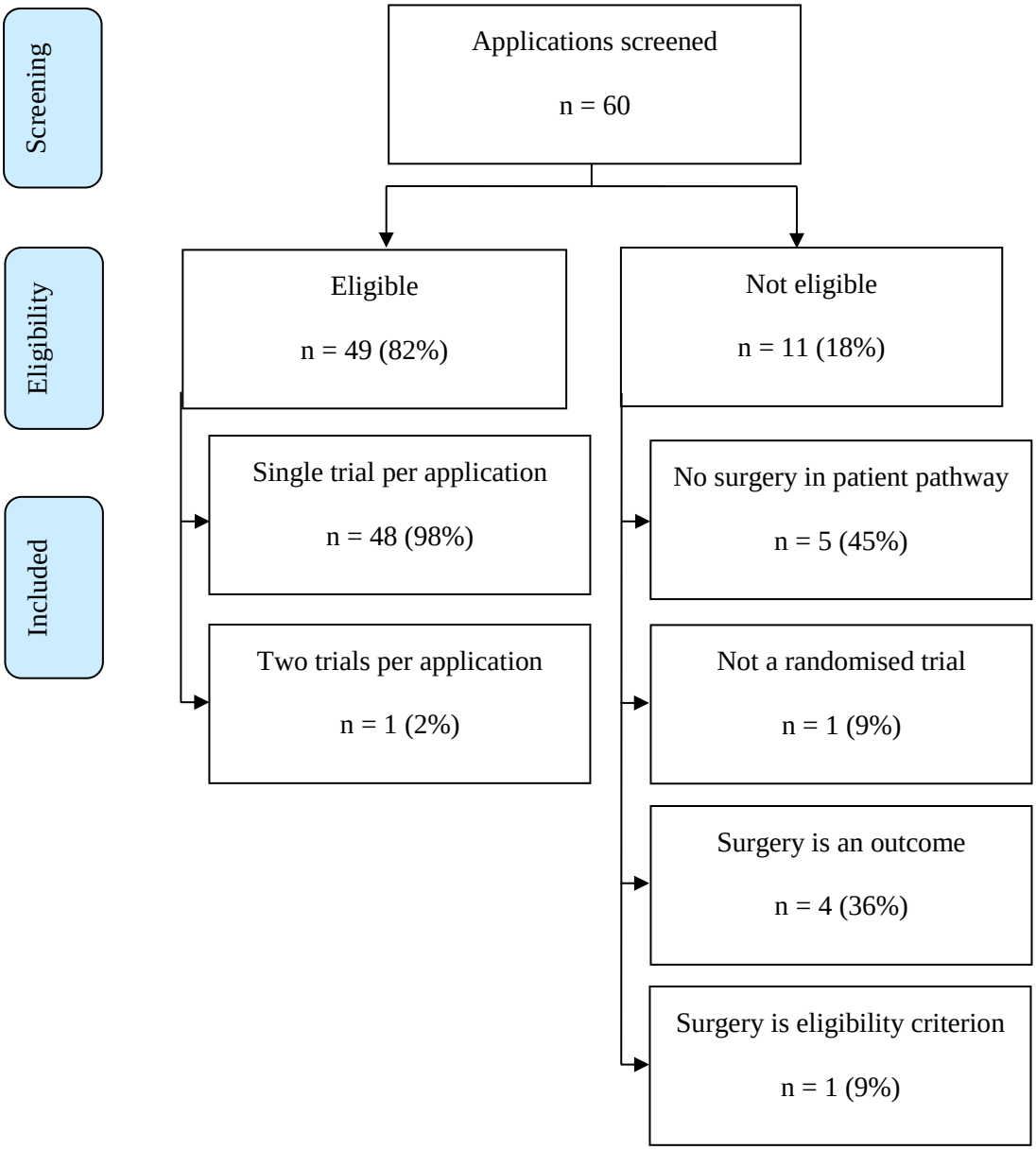
denoted by *[…]* and the addition of words or replaced words is denoted by *[words]* to aid understanding.

4.3. Results

4.3.1. Cohort details

The NETSCC compiled a report listing all surgery RCTs funded by the HTA and EME funding streams within the eligible period. Sixty potentially eligible studies were identified, of which 49 (82%) met the eligibility criteria following further central screening (*Figure 3*).

Figure 3: Flowchart of eligibility for grant applications



4.3.2. Double data extraction

Five articles were randomly selected from the eligible studies for double data extraction. Of 155 variables checked, two discrepancies were identified (1.3% error rate).

Table 10: Cohort summary

Item	Category	n	N	n/N%
Number of RCTs in application	One	48	49	98%
	Two	1	49	2%
Funder	HTA	44	49	90%
	EME	5	49	10%
Lead institution region	East	1	49	2%
	East Midlands	4	49	8%
	London	10	49	20%
	North East	7	49	14%
	North West	2	49	4%
	Scotland	10	49	20%
	South East	3	49	6%
	South West	4	49	8%
	Wales	2	49	4%
	West Midlands	4	49	8%
	Yorkshire and the Humber	2	49	4%
Trial start year	2012	3	49	6%
	2013	9	49	18%
	2014	26	49	53%
	2015	3	49	6%
	2016	1	49	2%
	2017	7	49	14%

Item	Category	n	N	n/N%
Source documents available [A]	Commissioning brief	15	49	31%
	Project description	40	49	82%
	Responses to board and peer review comments	40	49	82%
	Protocol	42	49	86%

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

4.3.3. Cohort summary

The majority of the applications were funded by the HTA (n=44/49, 89%) and had start dates from 2014 onwards (n=37/49, 76%); see *Table 10*.

Documents for review consisted of commissioning briefs (n=15/49, 31%), project descriptions (n=40/49, 82%), applicant responses to board and peer review comments (n=40/49, 82%) and protocols (n=42/49, 86%). Either the protocol or project description was available for all applications; see *Table 10*.

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

4.3.4. Trial demographics

Trials were primarily two-armed (n=45/50, 90%) and of a parallel design (n=49/50, 98%). Eight did not use a pilot or feasibility study (n=8/50, 16%). (41) In 11 studies (n=11/50, 22%), surgery was not the intervention of interest and delivered as part of the patient pathway. Where surgery was the intervention of interest (n=39/50, 78%), 21 compared against surgery, for example minimal access vs. open surgery (n=21/39, 54%). The remaining eighteen compared

surgery against a non-surgical comparator (medical comparator e.g. injection vs. surgery: n=7/39, other e.g. active monitoring and surgery vs. active monitoring only: n=11/39) (see *Table 11*).

Table 11: Trial design details

Item	Category	N	N	n/N%
Type	Parallel	49	50	98%
	Sequential (52)	1	50	2%
Number of trial arms	2	45	50	90%
	3	4	50	8%
	4	1	50	2%
Use of pilot or feasibility study, internal or external (41)	Both pilot and feasibility	2	50	4%
	Pilot only	29	50	58%
	Feasibility only	11	50	22%
	No	8	50	16%
Nature of surgery delivered	As an intervention	39	50	78%
	As part of patient pathway	11	50	22%
If intervention comparator	Surgery	21	39	54%
	Medical	7	39	18%
	Other	11	39	28%
If surgical comparator	Alternative surgical procedure	13	21	62%
	Change to a component of the same procedure	6	21	29%
	Same procedure delivered at a different time point	2	21	10%

4.3.5. Recruitment and randomisation

Patients were the randomisation unit in all trials and primarily allocated to equal groups (n=48/50, 96%). The majority stratified randomisation (n=46/50, 92%). In trials comparing two surgeries, there were no expertise-based designs. (42) *Table 12* provides more detail.

Table 12: Recruitment and randomisation

Item	Category	n	N	n/N%
Method of randomisation	Dynamic allocation	23	50	46%
	Minimisation	21	23	91%
	Other	2	23	9%
	Block	17	50	34%
	Not specified	10	50	20%
Allocation ratio	Equal	48	50	96%
	Unequal	1	50	2%
	Not specified	1	50	2%
Randomisation unit	Patient	50	50	100%
Randomisation stratified	Yes	46	50	92%
	No, not specified	4	50	8%
Multiple countries participating	Yes	3	50	6%
	No	45	50	90%
	Not reported	2	50	4%
If yes, stratified by country	Yes	1	3	33%
	No	2	3	66%
Multiple centres participating	Yes	49	50	98%
	No	1	50	2%
	Not reported	0	50	.
If yes, stratified by centre	Yes	28	49	57%

Item	Category	n	N	n/N%
	No, justification provided	1	49	2%
	No, by other variables	17	49	35%
	No, not stratified	3	49	6%
Multiple surgeons participating	Yes	22	50	44%
	No	0	50	.
	Not reported	28	50	56%
If yes, stratified by surgeon	Yes	8	22	36%
	No, justification provided	0	22	.
	No, by other variables	13	22	59%
	No, not stratified	1	22	5%
If yes, multi-centre	Yes	21	22	96%
	No	1	22	5%
If yes, stratified by	Centre and surgeon	2	21	10%
	Centre, not surgeon	11	21	52%
	Surgeon, not centre	6	21	29%
	Neither centre nor surgeon	2	21	10%

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

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

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

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

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

Three trials were international, of which one stratified randomisation according to UK, or non UK, centre. *Table 13* provides more detail of intervention approach by stratification type.

Table 13: Stratification factors in multi-centre and multi-surgeon trials by intervention type

		Stratified by centre					Stratified by surgeon				
		Multi-centre	Yes		No		Multi-surgeon	Yes		No	
Comparator of interest		**N**	**n**	**n/N%**	**n**	**n/N%**	**N**	**n**	**n/N%**	**n**	**n/N%**
Surgery vs. surgery	Different surgical interventions	13	5	38%	8	62%	6	4	67%	2	33%
	Different components of the same intervention	5	4	80%	1	20%	6	3	50%	3	50%
	Different time points of the same intervention	2	1	50%	1	50%	0	0	.	0	.
Surgery vs. medical		7	5	71%	2	29%	2	0	.	2	100%
Surgery vs. other		11	5	45%	6	55%	3	0	.	3	100
Surgery occurred but not intervention of interest		11	8	73%	3	27%	5	1	20%	4	80%

4.3.6.Surgeon and centre credentials

Centre and surgeon credentials, or inclusion criteria of those delivering the intervention, were provided in 41 (n=41/50, 82%) and 36 (n=36/50, 72%) trials, respectively (*Box 10* and *Box 11*). Most common centre credentials were case volume (n=20) and required fields of expertise within centre (n=13). Examples of surgeon credentials were grade or experience (n=16) and study specific training (n=13).

Box 10: Centre level credentials

Centre credential provided	41
Case volume	20 (48%)
Fields of expertise within centre	13 (32%)
Experience required without definition	9 (22%)
Experience required with definition	8 (20%)
Good recruiting reputation	8 (20%)
Experience required with definition	8 (20%)
Access to equipment required	7 (17%)
Centre to undertake trial specific training	2 (5%)
Demonstrated ability to participate	1 (2%)
Interest expressed in specific treatment	1 (2%)
Prior number of cases required	1 (2%)
Centre delivers one treatment only	1 (2%)

4.3.7. Trial outcomes related to learning and clustering

Forty-one applications explored outcomes that may reflect variability in centre or surgeon skill (82%, *Box 12*). Common outcomes were safety events (n=36); time to recovery from surgery (n=13) and operative time (n=6).

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

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

Box 11: Surgeon level credentials

Surgeon credentials provided	36
Grade or experience	16 (44%)
Study specific training	13 (36%)
Experience required without definition	8 (22%)
Oversight of supervision	7 (19%)
Prior number of cases	7 (19%)
Self assessed ability	7 (19%)
Equipoise	4 (11%)
Known to be good recruiters	3 (8%)
Case volume	2 (6%)
Local practice relevant	1 (3%)

4.3.8. Statistical considerations

Sample size adjustment

There were no examples of sample size adjustment for clustering at a centre level. Three applications adjusted the sample size for surgeon using an ICC and a fourth chose not to adjust although provided justification:

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

Exploratory analysis

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

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

Box 12: Outcomes

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

[A] Established using qualitative methods; [B] Feasibility outcome

Exploratory analyses considering differences by surgeon were planned in seven applications, of which three also explored by centre. Two analysed descriptively by surgeon grade and four via subgroup analysis: one modelled the learning curve using outcomes operation time and complications as a proxy to measure the task efficiency of the surgeon, one planned to explore

trends and changes over time between experienced and less experienced surgeons, one via a qualitative analysis and the final where patients were sampled for observations in theatre according to their treating surgeon's grade. As with centre, one application planned a sensitivity analysis that adjusted for surgeon.

Formal adjustment

Formal adjustment for multiple centre or surgeon effect was planned in 21 and 15 applications, respectively. *Table 14* provides more detail. When formally adjusting for centre, nine planned to use a random effect and thirteen did not specify. Similarly, six planned to adjust for surgeon using a random effect and nine did not specify. Of the applications planning a formal adjustment, 17 (n=17/21, 81%) of applications adjusting for centre and nine (n=9/15, 60%) adjusting for surgeon did so in addition to stratifying randomisation by these variables.

The two applications that planned to stratify by both centre and surgeon (*Table 12*), also planned formally adjusting analysis by these factors.

Table 14: Planned statistical adjustments through analysis in multi-centre and multi-surgeon trials

		Centre			**Surgeon**		
		n	**N**	**n/N%**	**n**	**N**	**n/N%**
Adjustment made		21	49	43%	15	22	68%
Type of effect	Fixed	0	21	.	0	15	.
	Random	9	21	43%	6	15	40%
	Time varying	0	21	.	0	15	.
	Not specified	12	21	57%	9	15	60%
Stratified and adjusted	Yes	17	21	81%	9	15	60%

4.3.9. Funder led considerations

Commissioning briefs

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

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

Changes driven by funder

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

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

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

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

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

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

4.4. Discussion

This review has investigated the decision-making behind intended design and analysis of 50 randomised surgical trials funded by the NIHR EME and NIHR HTA programmes from 2012 to 2016. These results show frequent consideration of centres and surgeon impact during design, and these may be funder led, due to concerns around homogeneity or generalisability of results. This review provides a cross sectional insight into current practice of researchers, and expectations of reviewers and funders, during trial design within two streams of a major UK funder. (49, 50)

The need for transparency around learning curves and clustering are highlighted within reporting of non-pharmacological interventions guidelines, (21, 30) and *Chapter 3* identified a deficiency in adherence to these. In contrast, this review identifies that considerations to manage learning and clustering are made, by both researchers and funders, during development of trials funded by a prestigious body. For example, 30% of multi-centre and 12% of multi-surgeon studies reported a statistical adjustment of these within published manuscripts. This was 43% and 68% respectively in this cohort. When randomisation was stratified by centre or surgeon, this was accounted for in the analysis in 30% of multi-centre and 40% of multi-surgeon trials in the published manuscripts, as oppose to 81% and 60% in this cohort. In drawing this comparison it is important to differentiate between the intended audiences. The detail required for a funding application, assessed by clinicians and methodologists/statisticians, may exceed that required to communicate results to a clinical audience. This demonstrates benefit in exploring unpublished trial documentation to understand approaches to trial design and analysis and highlights the need for improvements to transparent reporting.

The cohort included successful applications to the NIHR 2012 call for *Applied Health in Surgery*. (6) This call recognised the need to increase research-based evidence in surgery. Applications were invited that evaluated technology-driven implanted or implantable medical devices, surgical procedures or surgical services. As a clinical trial is typically a major financial investment, (53) applicants need to assure funders that their proposal is important, well designed and demonstrates scientific value to add to the current evidence base. Each application undergoes a peer review process, where 'experts' critically review the trial to ensure standards are met in terms of design, quality, feasibility, acceptability and importance of the topic. (21, 30) A strength of this review is the insight into the designs proposed to funders, and impact of feedback on subsequently funded studies.

Whilst the degree of learning and clustering will vary trial-to-trial, many interventions require surgical skill in their delivery regardless of whether or not the surgery is the intervention of interest. The impact of any potential imbalance in delivery on comparing interventions should be considered at trial outset routinely. Early and careful consideration will ensure that procedures are standardised where possible and data captured to support any further investigation. Such that, in severe cases, the trial team can alleviate any doubts about homogeneity raised by the medical community should the trial results be questioned. (16) These results indicate funder awareness of this early consideration, with one of the two examples of balancing randomisation by surgeon following funder recommendation in a trial where surgery was not the intervention of interest.

When interpreting these results, it is important to consider the limitations of this review. First, only successful applications could be included due to confidentiality constraints. It is therefore not possible to determine whether the management of learning and clustering contributes to the success of the application. However, given that the application process consists of iterations whereby peer reviewers are able to request that researchers address paucities in their

application, it is unlikely that a promising application, lacking in the appropriate considerations, would be deemed unsuitable for funding outright. More likely, researchers would be given the opportunity to make these considerations during this iteration process. Second, as part of this iterative review, it is possible that additional discussions at the funder board meetings did not make it in to the comments fed back to applicants. This could mean that funders raised these issues more frequently than this review suggests. Third, due to the nature of the grant application process, the funder impact observed may be in part due to an increased awareness of the reviewers involved. Fourth, this work has focussed on a single funding body that primarily supports UK based research. However, trials supported span a wide range of surgical specialties and health care conditions and results from this review will be generalisable to other funding bodies with a similar peer review process. Fifth, this review is cross-sectional. Since its undertaking, many advances in guidelines have been developed, see *Chapter 2*, therefore the results will likely be different if repeated on a more recent sample.

4.5. Conclusions

Fundamental to trial design and analysis is understanding the objectives. While considerations relating to clustering and learning effects are not widely reported in main trial publications, these results indicate both funders and researchers consider these aspects in order to address a specific research question. Such issues may have varying relevance depending on the overall design of the trial. A very pragmatic study may deliberately include surgeons and centres of all types and have less emphasis on expertise and learning, whereas the delivery of the intervention in more explanatory studies is critical and requires consideration during design and analysis. Another approach to overcoming these issues is to provide quality assurance of the intervention. Early work to develop methods to achieve this exist and it is expected that this will expand in the future. (54) What is important is that variability about treatment delivery is understood, care is taken when defining the interventions per protocol and that

areas where issues may arise are identified early in the trial so that they can be integrated, as appropriate, into trial design.

Furthermore, these results provide insight into the promising role of the funder as a driver to improving the, long criticised, surgical evidence base. The funder, who has influence over whether or not and how studies are carried out and has been suggested as a driver for improving the quality of research during the period of growth for surgical trials, (55) can play a valuable role in ensuring that future trials do not have the same shortfalls as those in the past.

Chapter 5 : Views and current practice within clinical trials units

5.1. Introduction

Chapter 3 identified a lack of adherence to reporting guidelines (21, 30) and *Chapter 4* identified the funder as a driver for improving the surgical evidence base for future trials. To complete understanding of potential barriers to improvement, with regards to the methodological and statistical considerations for surgical trials, statistical representatives from Clinical Trials Units were surveyed.

This chapter begins by providing further justification for this survey (*Section 5.1*). *Section 5.2* described in detail the methods used and subsequent results are provided in *Section 5.3.*

The work in this chapter has been published in Trials and I am first author. (56)

5.1.1. Variation in approaches to management

Whilst the notion of clustering and learning is familiar to many statisticians and methodologists, 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 multi-centre trials was undertaken within the UK Clinical Research Collaborative registered Clinical Trials Units. (57)

5.1.2. Aims

This survey aimed to ascertain UK wide experience of running multi-centre 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.

5.2. Methods

5.2.1. Survey delivery

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

5.2.2. Survey development

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 and is shown in *Appendix Box 2*.

This survey was developed to establish experience in multi-centre 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 CG, EJC, GB, JAC and JMB. ((12, 13, 16), *Box 13*). 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.

Box 13: Example trial scenarios

	Scenarios
A	A trial with a large (each centre recruiting at least ten patients) sample size, recruiting in several centres each with multiple treatment providers.
B	A trial with a small (each centre recruiting 2-3 patients) 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.

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

5.3. Results

5.3.1.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). *Table 15* 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.

Table 15: Participation summary

	Number of Units		
Participation status	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%

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%). *Table 16* provides further detail.

Units listed on the UKCRC Resource Finder (57) as conducting cluster or surgical trials had participation rates 94% (n=16/17) and 92% (n=33/36) respectively (*Table 17*). Units with a methodological research area in complex interventions participated with a rate of 90% (n=35/39).

Table 16: Role of Unit representative

	Number of Units		
Role	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%

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*, *Table 18*).

5.3.2.Managing effects through design

4.3.2.1. Clustering

Twenty-five Units had undertaken multicentre trials that did not stratify by centre (n=25/44, 57%, *Question 2*, *Table 19* and *Table 20*). 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 the Unit's 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 13*, *Table 19* and *Appendix Table 2*), 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 in *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, Appendix Table 2*). Two-thirds (n=8/12, 67%) commented on the feasibility of stratifying by treatment provider. Reasons were as follows: 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]; and 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 19)*, nineteen Units used the same approach across all scenarios in which 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.

4.3.2.1. 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 19*). 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 [ID15] 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, *Question 5*, *Table 19*).

Table 17: Completion rates by speciality trial types

	Specialises in trial of type											
	Complex intervention				Surgical intervention				Cluster randomised			
	Yes		No		Yes		No		Yes		No	
N	39		11		36		14		17		33	
Completer status	**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%

Table 18: 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%

Table 19: Managing learning and clustering by design

			Response statistics		
Question		**Category**	**n**	**N**	**n/N%**
2	Does your unit have any multi-centre trials that do not stratify randomisation by centre? *See Table 20 for further details.*				
		Yes	25	44	57%
		No	18	44	41%
		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 Appendix Table 2 for further details.*				
a	Large sample size, [A] recruiting in several centres, each with multiple treatment providers				
		Experience in trial type	39	44	89%

Question		Response statistics		
	Category	**n**	**N**	**n/N%**
	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, [B] 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%

Question		Response statistics		
	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%
	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%

Question		Response statistics		
	Category	**n**	**N**	**n/N%**
	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%

Question		Response statistics		
	Category	n	N	n/N%
	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 [C]	13	44	30%
	No, with justification	1	44	2%
	No	26	44	59%
	No response	4	44	9%

[A] With centres each recruiting at least ten patients.

[B] With centres each recruiting 2-3 patients.

[C] 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 20: Reasons for having multi-centre studies that do/do not stratify by centre (Question 2)

Unit has multi-centre trials that do not stratify randomisation by centre?	**Yes**		**No**	
N	25		18	
Reason(s) provided	**n**	**n/N%**	**n**	**n/N%**
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. (58)	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	0	0%	1	6%

balance of resources and avoid confounding. There is a lot of academic debate. See Torgerson.				
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%

5.3.3.Managing effects through analysis

4.3.3.1. 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 21* and *Box 14*). 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]; centre 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 21* and *Appendix 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." [ID15]

In trials stratified by treatment provider, 36% also subsequently adjusted the analysis (n=16/44, 36%, *Question 6, Table 21* and *Box 14*). 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 21* and *Appendix 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."* [ID15]

When responders were asked to revisit the scenarios in *Box 13,* this time to consider investigating treatment by centre or treatment provider (*Question 9, Table 21*), 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, Appendix Table 3*). General themes for additional information provided were as follows: 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 21*), 24 Units used the same approach across all scenarios and twelve utilised a scenario specific approach (same: n=24/44, 55%; 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 explored heterogeneity by centre when a positive treatment effect is found (n=32/44, 73%, *Question 10a, Table 21*), whereas fewer explored heterogeneity

by treatment provider (n=12/44, 27%, *Question 10b*, *Table 21*). 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). *Appendix Table 4* and *Appendix Table 5* provides further detail.

4.3.3.1. 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 21*), 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], with one specifying:

"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].

Table 21: Managing learning and clustering by analysis

Question			**Response statistics**		
		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.			

Question		Response statistics		
	Category	**n**	**N**	**n/N%**
	See Box 14 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? *See Box 14 for further details.*			
	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%
	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?			

Question		Response statistics		
	Category	**n**	**N**	**n/N%**
	Yes	39	44	89%
	But only if it was used to stratify randomisation	18	39	46%
	Always	6	39	15%
	Sometimes [A]	15	39	38%
	No, never	3	44	7%
	No response [B]	2	44	5%
a	If yes, and assuming that the sample size allows either, would you treat this effect as fixed or random? *See Appendix 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? *See Appendix Box 4 for further details.*			
	Yes	26	44	59%
	But only if it was used to stratify randomisation	8	26	31%
	Always	0	26	0%
	Sometimes [C]	18	26	69%
	No, never	13	46	30%
	No response [D]	5	44	11%

Question			Response statistics		
		Category	**n**	**N**	**n/N%**
	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 Appendix Table 3 for further details.*			
	a	Large sample size, [E] recruiting in several centres, each with multiple treatment providers			
		Experience in trials 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%

Question		Response statistics		
	Category	**n**	**N**	**n/N%**
b	Small sample size, [F] recruiting in several centres, each with multiple treatment providers			
	Experience in trials 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 trials 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 trials type	21	44	48%
	Centre	5	19	24%
	Treatment provider	3	19	14%

Question			Response statistics		
		Category	**n**	**N**	**n/N%**
		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 trials 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 Appendix Table 4 for further details.*			
		Yes	32	44	73%
		No	9	44	20%

Question	Category		Response statistics		
			n	**N**	**n/N%**
	No response		3	44	7%
	i.	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 Appendix Table 5 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, do you explore by analytical methods e.g. significance testing?			

Question		Response statistics		
	Category	n	N	n/N%
	Yes	9	12	75%
	No	1	12	8%
	No response	2	12	17%

[A] "Sometimes" here is "usually" – it is a rare exception where we don't. [ID10].

[B] No Standard Operating Procedure in place. [ID3].

[C] "Sometimes" here is "usually" – it is a rare exception where we don't. [ID10].

[D] No experience in trials of this type. [ID1] Not applicable. [ID2].

[E] With centres each recruiting at least ten patients.

[F] 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].

Box 14: Other grouping rules when randomisation is stratified by (a) centres or (b) treatment providers (Question 6)

Centre stratified (n=24)	
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 SAP 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 (n=17)	
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 (n=16)	
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 SAP 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 CRF 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.

5.4. Discussion

This survey identifies that, despite multi-centre trials being prominent across all Units, there is 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, and 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 multi-centre trials, although the 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 multi-centre trials investigating varying types of interventions. Whilst acknowledging that different approaches may be more suitable to different trial types, they indicate a 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. 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, and guidance for practical issues like these are available. (59) 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 (16) was most common, with almost all responders (89%) applying this approach to studies within their Unit. Less than one third indicated experience in conducting expertise-based designs, a design that can be particularly useful when comparing substantially different interventions. This finding suggests that these designs are more commonly implemented than suggested by the literature. (25, 26, 60) 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 recommendations being explicitly incorporated into the survey questions. (*Appendix Box 2*) (14, 16) Additional documents within the International Council for Harmonisation (ICH) Series provide further guidance beyond *ICH E9*. (16, 28, 33, 34) The Consolidated Standards of Reporting Trials (CONSORT) statement and relevant extensions provide direction valuable at study design despite the document being developed to support reporting. (21, 30) Further, there have been further update to the MRC Guidelines on Complex Interventions, providing further options for trial design, see *Chapter 2* for further details. (13, 61) 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. (62) However, the ability to identify and explore heterogeneity at the analysis stage is important for generalisability for all trials.

Strengths of this investigation are 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. (63) 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 (28) 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, leaving global practice unknown. However, the survey drew upon internationally accepted guidelines (28) 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 many pragmatic large-scale trials do not have 'learning' effects because they include interventions that are stabilised and already 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, and one responder per Unit may result in experiences reported for larger Units not being indicative of all studies run. However, responders were able to complete the survey with additional support within their Unit if required.

5.5. Conclusions

This survey is the first to report on the experience and management approaches with regards to clustering effects and the learning curve in multi-centre 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 RCTs was developed more generally to support consistency in approaches across a more conventional RCT, (16, 28, 34) with the development of more intervention specific guidelines being established following these to address the additional complexities across different types of trials. (13, 21, 23, 30) 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.

Chapter 6 : Example trials involving surgery

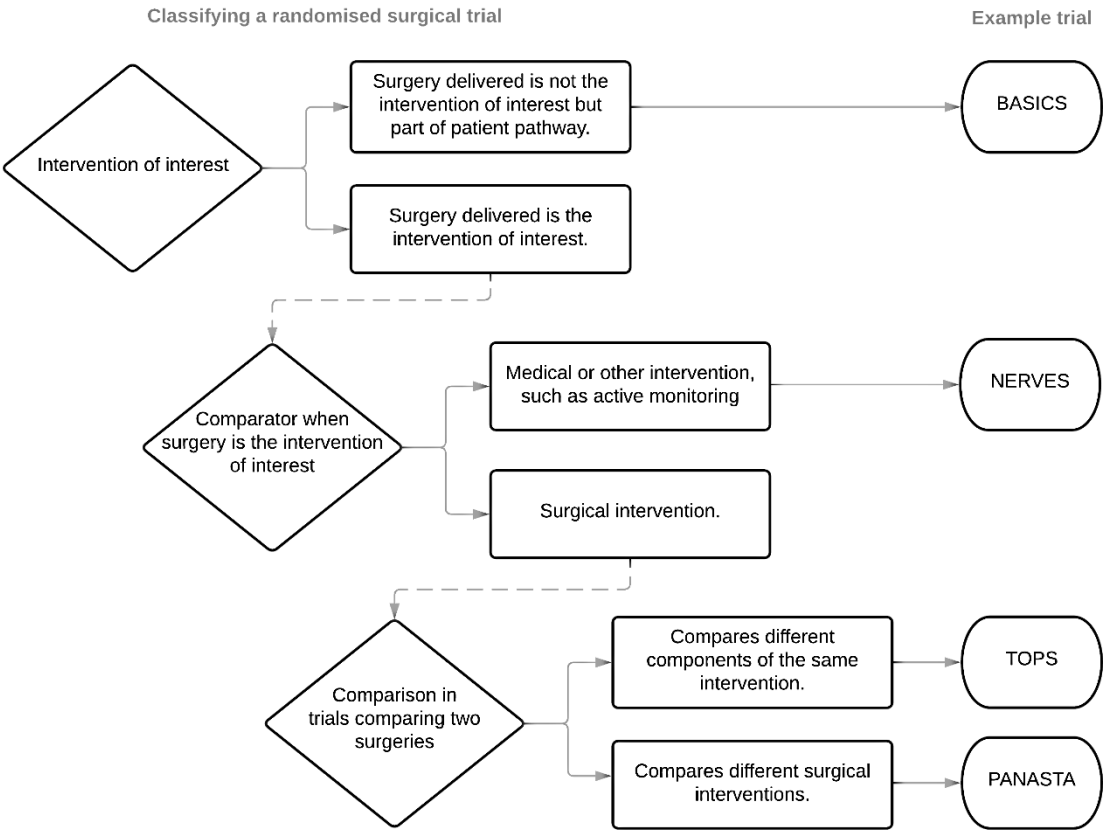
6.1. Introduction

This chapter presents real trial examples to highlight when and how consideration of learning and clustering effects may be appropriate. Examples have been purposely selected from Liverpool Clinical Trials Centre portfolio such that they represent different types of surgical trial. Their summary will demonstrate one Unit's experience in approaches that can be taken to manage learning and clustering in practice, thereby aiding the discussion around these considerations in pragmatic research in a surgical setting. A number of these trials will be in later chapters illustrate the application of analytical methods that adjust for learning and clustering.

6.2. Classifying surgical trials

The reviews of the trials within the literature or grant applications conducted in *Chapter 3* and *Chapter 4* classified trials according to the type of interventions that they compare, see *Figure 4*. Available guidelines for complex or surgical trials, summarised in *Chapter 2*, tend to focus on their applicability and relevance in relation to the interventions. However, issues relating to differential expertise bias apply to any trial involving an invasive procedure. (64) When surgery is delivered within the trial but is not the intervention of interest, the level of clustering or learning curve associated may not be considered as strong, but it is still possible that poor design, or failure to consider differential expertise bias, may lead to any differences observed in treatment effects being wrongly attributed to the treatment itself and not differences due to performance bias. For this reason, trials that involve an invasive procedure, where surgery is delivered as part of the patient pathway, are included within this summary.

Figure 4: Defining type of surgical trial for examples



6.3. Case studies

All trials were designed and analysed by the Liverpool Clinical Trials Centre – a UK Clinical Research Collaboration (UKCRC) registered Clinical Trial Unit. (65) The Centre has a specific research stream for surgical interventions and has 14 trials in their portfolio at the time of writing. All trials presented within this chapter are pragmatic multi-centre and multi-surgeon thereby representing up to three levels of nesting (patients within surgeon within centre).

A case-based approach was adopted to present the approaches used to manage learning and clustering at the design and analysis stage. Four surgical trials were purposefully selected to demonstrate learning and clustering such that they represent different surgical trial settings. Each differing in terms of the nature of their surgical comparisons, see *Figure 4*.

Examples presented also vary by surgical area, sample size and differing levels of approaches used to account for learning and clustering at the design and analysis stage. A summary for each trial is provided including intervention type, setting, patient population, forms of treatment and outcomes followed by a description of the approaches used at the design and analysis stage to manage the potential effect of learning and clustering.

6.4. Example trials

6.4.1. A trial where surgery is delivered as part of patient pathway: the BASICS trial

Ventriculoperitoneal shunts (VPS) as a treatment for hydrocephalus is one of the most common procedures performed in neurosurgical units. (66) BASICS was a parallel, three-arm, blinded, multi-centre randomised controlled trial comparing antibiotic-impregnated, silver-impregnated and standard (non-impregnated) VPS in patients with hydrocephalus undergoing insertion of their first permanent shunt. The trial population included 1605 patients (children and adults) with hydrocephalus of any aetiology requiring first shunt. Patients were recruited from 19 neurosurgical wards across the United Kingdom and Ireland. Randomisation (1:1:1) was stratified by centre and operating theatre within centre, for logistic reasons to allow the randomisation envelopes to be accessed when required, and undertaken by the operating surgeon within theatre at the time of insertion. All shunts used in the trial were CE marked medical devices being used for their intended purpose. The primary outcome was time to shunt removal for shunt infection. Secondary outcomes included time to shunt failure for any cause, reason of failure (infection, mechanical), and time to removal for shunt infection following a clean revision.

This trial has a published protocol and main results paper that can be accessed for further information. (67-69)

Considering learning

BASICS was a device trial where the intervention of interest was different coatings of the shunt inserted. The surgical technique across the three arms of the study, insertion of a shunt, was the same, although the devices themselves varied in their size and flexibility. At the time of trial design, feasibility data demonstrated that antibiotic and standard shunts were most commonly used.

The key differences between the shunts are in their length and flexibility. Silver shunts are more rigid than other types. During insertion, a metal rod is used to prevent the more flexible shunts from bending, once in situ this rod is removed. The silver shunt is strong enough and does not require a rod to assist with insertion. There may therefore be a surgeon learning curve present for surgeons to overcome initial difficulties with the size and flexibility of the device, despite the surgical technique across the three arms of the study being the same.

During the trial there was some evidence of clinician preference regarding shunt selection which led to exclusion of 17% (n/N=74/435) eligible patients during the screening process. Further details of the preference were not collected and only a few consultants willingly offered details with one preferring the antibiotic shunt and two preferring the standard. Once randomised, 1% (n/N=16/1601) of patients did not received the shunt allocated with reasons for not adhering to randomisation occasionally specific to the shunt allocated.

Considering clustering

Age was a known prognostic indicator of shunt failure, and most recruiting centres treated paediatrics or adults, rarely both. It was expected therefore that centre would be a prognostic indicator of outcome. Important considerations for design and analysis to allow for this included balancing randomisation with respect to centre through stratification and subsequently adjusting the analysis for age of recruiting hospital. The results supported this design consideration, with one quarter of the randomised patients being under one year of age

and results confirming a greater risk of revision, and infection, within the paediatric cohort. However, there may be further differences within the groups that this design and analysis approach did not capture. Variation of shunt size would typically be more of a concern within very young patients and the difference between a one year old and a 15 year old, who would fall within the same strata, will be greater than a 15 year old and a 19 year old who would be categorised in separate risk strata.

Further, BASICS was a pragmatic study and did not require participating centres to change their routine shunt insertion practice. As a relatively common treatment within neurosurgery there is potential for a wide surgical experience base and variations to routine pre and post-operative care. Therefore, an expertise and practice imbalance with regards to shunt insertion between arms, by chance, was possible.

Managing by design

Clustering, by centre, was managed through stratification of the randomisation schedule. Stratifying by recruiting centre ensured balance within centre of the three shunt types.

The trial team required that recruiting centres met the BASICS Centre Inclusion Criteria, which required aspects relevant to clustering that centres were regional neurosurgical units treating adults or paediatrics and provided evidence that they were able to recruit a minimum of three participants per month, which demonstrates that the centre has sufficient experience in treating eligible patients. No criterion covered aspects of learning in terms of expertise in fitting each shunt type. This was because differences with regards to shunt variations were raised during the trial and not at the design stage and meant that trial specific surgical training, or quality assurance monitoring, was incorporated in the design. This highlights a need for triallists to discuss in detail any differences in the treatments and how they are delivered early in the study, even in trials where surgery is not the intervention of interest, as in the BASICS trial.

Managing by analysis

No analysis to explore learning was undertaken as part of the main trial analysis. However, to facilitate any analysis of learning and clustering, data collection forms were designed to capture information that would allow heterogeneity by centre, or surgeon, to be explored analytically if required. For example, details of the lead and other members of the operating team were collected, including as grade of operator. Data were also collected on patient level factors that may indicate a greater risk of shunt failure, such as whether the patient was pre-term at birth for paediatrics.

Analyses of the primary outcome, and selected secondary outcomes, were adjusted for age of demographic within recruiting hospital, for example paediatric, due to the prognostic value of age. Adjusting for both values was not possible due to dependency of age group on recruiting centre. Instead, heterogeneity in treatment between centres was explored graphically and by summary statistics. Infection rates (primary outcome), revision rates (secondary outcome) and confidence intervals around these were reported, split by shunt type, on a per centre level. No analysis to explore clustering by surgeon was undertaken and this is further explored within *Chapter 7*.

6.4.2. A trial comparing a surgical to a non-surgical intervention: the NERVES trial

Sciatica is the name given to any pain caused by irritation or compression of the sciatic nerve and affects approximately 3% of the UK population at any one time. (70) NERVES compared two treatments for sciatica: injection and spinal surgery. NERVES was a parallel, national, two-arm, open-label, multi-centre randomised controlled trial comparing the clinical and cost effectiveness of trans-foraminal epidural steroid injection to surgical microdiscectomy for the treatment of chronic radicular pain secondary to prolapsed intervertebral disc herniation. The trial population included 163 patients with sciatic pain endured for between six weeks and

twelve months whose symptoms had not been improved by at least one form of conservative (non-operative) treatment. Patients were recruited from eleven outpatient neurosurgical, pain and orthopaedic clinics in the United Kingdom. Randomisation (1:1) was stratified by centre. The primary outcome was patient-reported disability measured using the Oswestry Disability Questionnaire scale at 18 weeks post randomisation. Secondary outcomes included disability and pain scales using numerical pain ratings, modified Roland-Morris and Core Outcome Measures Index, at 12-weekly intervals.

This trial has a published protocol and main results paper that can be accessed for further information. (70-72)

Considering learning

As the trial compared a surgical intervention to a non-surgical intervention, treatment provider specialty could differ between arms. The injection comprises a local anaesthetic to numb the sciatic nerve, providing short-term pain relief, while the steroid has a long-term effect reducing inflammation in the joint and around the nerve. Spinal surgery removes the part of the disc that is causing the sciatica. Prior to the trial, both interventions were used routinely and a learning curve was not deemed to be a concern. If it were, then the triallists would likely need to allow for different learning curves per treatment arm, due to the substantially different nature of the treatments.

Considering clustering

NERVES is an example of a trial where the treatment provider delivering the intervention treats patients in one arm only, or is more likely to offer one treatment than the other. For example, an anaesthetist would only treat patients with the injection, and an orthopaedist may deliver both. In this case, the probability of a patient being assigned to an anaesthetist cluster and an orthopaedic surgeon cluster is not equal, as the anaesthetist can only conduct one of

the interventions. If the treatment groups are assigned to clusters with different probabilities, then the correlation between treatment assignments is non-zero. The ICC is generally not known prior to the trial commencing and, unless there is evidence to the contrary, should always be assumed to be non-zero. When the correlation between treatment assignment and the ICC are both non-zero then clustering should, at design, be treated as non-ignorable clustering. (73)

However, in NERVES, the sample size prevented further analysis of clustering. The sample size was 163 patients, with half the patients being recruited from a single centre and the remaining ten centres recruiting a minimum of one patient and a maximum of 14. This confined the amount of meaningful analysis that could be performed at the centre level, and treatment provider level due to their nesting within centre, and presents a common issue in trials. NERVES therefore highlight that sample size per centre, or treatment provider, should be considered at design when triallists consider clustering, or learning, a particular interest.

Managing by design

Clustering, by centre, was managed through stratification of the randomisation schedule. Stratifying by recruiting centre ensured balance within centre of the two treatment arms.

The trial team required that recruiting centres met the NERVES Centre Inclusion Criteria. Criteria relevant to clustering were that centres were neurosurgical, pain and orthopaedic clinics that receive patients from tertiary referral centres (General Practitioners, allied health professionals and non-spinal consultants).

Surgeons and anaesthetists were selected according to local practice and pre-requisites for treatment providers were arm specific. For the injection, treating specialists could include pain specialists, radiologists, anaesthetists and surgeons. As part of the inclusion criteria, treatment providers were required to perform the injection as per the trial protocol to minimise

variability in treatment delivery in the injection arm, with aspects such as the dosage and type of steroid for use being compulsory. For the surgery, treatment specialists were orthopaedic or neurosurgical consultants, or equivalent e.g. associate specialist. Deliverers could also be a specialist trainee directly supervised by a consultant. The surgical arm was performed as per local policy within a treatment window defined within the protocol.

Managing by analysis

No analysis to explore learning was undertaken. However, data collection forms were designed to capture information that would facilitate any analysis of learning and clustering if required. For example, the treatment provider's name, specialty e.g. orthopaedic and level e.g. trainee were collected, as were details of the wider team delivering the treatment. Data were also collected on patient level characteristics that may impact treatment success, such as duration of symptoms.

Analyses of the primary outcome, and secondary efficacy outcomes, were adjusted for recruiting centre. Centre was treated as a random effect and the outcomes compared using a linear mixed effect model, with fixed effects baseline score and treatment group. No analysis to explore clustering by treatment provider was undertaken.

6.4.3. A trial comparing different components of the same intervention: the TOPS trial

Clefts of the lip and/or palate are among the most common birth anomalies, occurring with an incidence of 1 in 600 births. (74) The timing of palatal surgery has been a controversial issue since the 1930s. (75) TOPS was a parallel, international two-arm, assessor-blinded, multi-centre randomised controlled trial comparing primary surgery, using the Sommerlad technique, for cleft palate at six months or twelve months of age (corrected for gestational age). The trial population included 558 infants with a diagnosis of non-syndromic isolated cleft palate who were considered medically fit for operation at six months, corrected for

gestational age. Patients were recruited from 23 centres in Brazil, Denmark, Norway, Sweden and the United Kingdom. Randomisation (1:1) was undertaken using a minimisation program within strata defined by surgeon and size of cleft (soft palate only or soft and hard palate). The primary outcome was insufficient velopharyngeal function at five years of age. Secondary outcomes, measured across twelve months, three years and five years of age included growth, safety of the procedure, dentofacial development, speech, hearing and middle ear function.

This trial has a published protocol and statistical analysis plan that can be accessed for further information. (74, 76) The main trial results paper will be published in due course.

Considering learning

It is important to consider the experience of the treatment provider in any trial. TOPS presents an example trial where differential expertise is likely to exist within patient subgroups. TOPS was a surgical trial where the intervention, primary palatal surgery using the Sommerlad technique, (77) delivered across both arms was the same, with the patient age at primary surgery varying between aged six and twelve months.

The opening of the infant's mouth is smaller in younger children which can make fitting the tools into the mouth to conduct the operation more difficult. Infants who are younger at the time of the operation may therefore be more prone to scarring or surgical complications. Timing of surgery was known to vary widely across participating centres, from six months of age to eighteen months of age. No centre had experience in both surgical timings of interest. Surgeons within centres who routinely operated on older infants may struggle with the younger infants in the study. Similarly, those who are familiar with operating on younger infants may find operating on older children in the twelve months arm, where the opening for the surgery is larger, easier to undertake.

In addition to concerns around expertise in terms of the size of the child, one centre had no experience of the Sommerlad technique prior to trial participation. The independent data safety monitoring board and the trial management group therefore raised concerns of a learning curve in terms of the intervention.

Considering clustering

During the trial, differences between centres in terms of randomisation and adherence to timing, which may have stemmed from concerns around expertise, were noted. One-third of all reported ineligible infants were excluded as not medically fit for operation at six months, with this proportion varying from 0.0% to 48.6% across the recruiting centres.

Due to the trial recruiting on an international level, TOPS provides an example of a trial where there is good reason to expect clustering by centre. Differences in outcomes at the centres or surgeon level may also be impacted by environmental, socio-economic or treatment delivery differences, such as routine pre and post-operative care or availability of resources to support infants with additional needs.

Managing by design

Each participating centre was required to demonstrate set criteria prior to participating in the trial. These criteria included high volume of patients which demonstrates surgical experience, albeit at a centre level and not surgeon. The cleft team at each centre were required to comprise: cleft surgeon(s), nursing staff, cleft speech and language therapist(s), clinical geneticist/paediatrician, audiologist(s), orthodontist and social worker.

Prior to participation, a formal process of surgical standardisation took place for all participating surgeons. A designated surgeon, who developed the technique, acted as lead surgeon for surgical calibration and provided instruction via written text, video demonstrations

and illustrated seminar and discussion sessions. This instruction also included a calibration session in the operating room. Calibrated surgeons then completed the TOPS Surgical Calibration Training Log, this log was signed by the training surgeon to confirm the competency of the listed individual was competent.

Clustering, by surgeon, was managed through minimising randomisation on operating surgeon. There were a maximum two surgeons per recruiting centre. Minimising on surgeon ensured balance within surgeon in terms of surgical timing. Balance is important to achieve in this case due to the differential expertise in treatment arms, with no participating surgeon having experience in both surgical timings prior to the trial.

Managing by analysis

No exploration of learning was undertaken as part of the pre-planned main trial analysis. However, data were captured to facilitate analysis of learning and clustering. Centre policy of primary surgery age prior to involvement in the trial, details of the operator and patient level characteristics, such as age at primary surgery, were collected.

Differences in the primary outcome due to heterogeneity in treatment between surgeons were accounted for using a multilevel logistic regression model including: a random effect for surgeon and fixed effects for minimisation factor size of cleft at baseline, treatment allocation, and an intercept. This was a sensitivity analysis, adjustments were not made for the main outcome comparison. Due to the overlap between centre and surgeon i.e. maximum of two surgeons per recruiting centre no analysis to explore clustering by centre was undertaken.

6.4.4. A trial comparing different surgical interventions: the PANASTA trial

Pancreato-duodenectomy as a procedure is 100 years old. During the 1940s, it was refined and standardised (78) and is currently the most common operation used to treat pancreatic cancer.

(79) PANasta is a parallel, national, two-arm, double-blinded, multi-centre RCT comparing Cattell-Warren and Blumgart techniques of pancreatico-jejunostomy following pancreato-duodenectomy for presumed malignancy. The trial population included 238 patients, over the age of 18, undergoing an elective pancreato-duodenectomy for presumed malignancy. Patients were recruited from eleven centres in the United Kingdom. Randomisation (1:1) was stratified by pancreatic texture (soft or normal/hard), pancreatic duct diameter (normal (≤3mm) or dilated (>3mm)), and centre. The primary outcome was the presence or absence of a post-operative pancreatic fistula (POPF) up to 3 months post-surgery. Secondary outcomes included adjuvant therapy, mortality rate, overall survival, rate of delayed gastric emptying, rate of wound infection, rate of pulmonary infection, rate of intra and post-operative bleeding, rate of reoperation, length of hospital stay, and operation time. Secondary outcomes are assessed up to a maximum 12 months post-surgery.

The PANasta trial is the first multi-centre RCT comparing two types of duct-to-mucosa pancreatic anastomosis with surgical quality assurance. This trial has a published protocol and main results paper that can be accessed for further information. (78, 80)

Considering learning

PANasta is a trial assessing two different surgical techniques of pancreatico-jejunostomy (reconstruction of the pancreatic remnant) following pacreato-duodenectomy: Cattell-Warren and Blumgart. (81) During design, there was a lack of well-designed studies, with existing studies lacking surgical quality assurance and consideration of the surgical learning curve. (78) This led to the potential for the personal preference of surgeon and differing evidence base introduced the potential for the treatments being delivered varying by surgeon and centre. Ensuring that the techniques being compared were as standardised as possible and minimising differential expertise was a particular challenge for the trial team.

Considering clustering

Despite the delivery of the intervention being standardised as far as possible, adjuvant therapies were not controlled as part of the study and considered to be of importance by the research team. Generally, following the surgery the patient is assigned to chemotherapy, but the time to this adjuvant treatment, and more importantly whether or not the patient is left for longer to ensure that they are fit enough for chemotherapy, can vary centre to centre.

Managing by design

Centre Inclusion Criteria were defined to ensure sufficient experience using both techniques. Centres were tertiary pancreatic surgery referral units, hospital inpatient and outpatient units to ensure a more standard approach to treatment delivery and a similar case mix of recruited patients. The trial team required local expertise in pancreatico-jejenostomy methods of recruiting the pancreatic remnant with local Principal Investigators required to be familiar with both techniques.

A number of steps were taken to ensure standardisation of both techniques were undertaken. Consensus meetings identified key steps of each anastomosis, such as management of drains, which was then developed into a pilot phase operative manual. Investigator meetings were held to provide trial specific training. An operative manual was developed by testing the pilot phase operative manual in two centres for six months as part of an internal pilot. The pilot manual was reviewed and adapted where necessary prior to being finalised and rolled out amongst all centres. The manual defined steps for the anastomosis that were mandatory, prohibited and flexible. Finally, pre and post-operative photographs were taken collected to ensure surgical quality and consistency through assessment by the Chief Investigator and a second reviewer.

Clustering, by centre, was managed through stratification of the randomisation schedule. Stratifying by recruiting centre ensured balance of the two treatment arms within centre.

Managing by analysis

No analysis to explore learning was undertaken. However, details of the surgeon and patient characteristics were collected to facilitate analysis as required.

A sensitivity analysis was performed using the Cochran-Mantel-Haenszel test calculated across centres. Logistic regression was also used to investigate individual variation in outcomes by centre.

Differences in time to adjuvant treatment, and how this varied by centre, were explored as part of a post hoc analysis.

6.5. Discussion

The surgical trials presented provide real examples of trials where learning or clustering considerations were warranted and demonstrated how, through design and analysis, these effects were managed. All of the trials here provide real examples of the considerations and recommendations for design and analysis presented for surgical learning and clustering in *Box 6*, *Box 7*, *Box 8* and *Box 9* in *Chapter 3* being applied in practice.

The trials presented differ in terms of their interventions, presenting a traditional surgical trial, such as PANasta, alongside a trial where the surgery is delivered as part of the patient pathway but is not the intervention of interest as in BASICS, and highlights the overlap between how considerations, with regards to learning and clustering, can be addressed through design and analysis. The selection includes studies varying in patient populations, including trials recruiting adults only (PANasta), children only (TOPS), and all ages (BASICS). The importance of age across a number of trials is highlighted, such as the size of the child at the time of the operation in TOPS and the importance of age as a prognostic factor relative to the

outcome in BASICS. Conditions presented can be chronic, as in NERVES, and vary across surgical specialty. The outcomes across all of the trials vary, yet in most cases the impact of centre or treatment provider, on these warrants consideration when designing the trial, with all studies aiming to standardise delivery using various methods.

Real examples of considering the learning curve are presented, and demonstrate how the recommendations, presented in *Box 7* in *Chapter 3*, can be followed. All trials in this chapter defined, to some level, a minimum experience required to deliver the intervention at centre or surgeon level (*Recommendation LC-1*). Examples included defining the specialty field of the deliverer, and this could be arm specific as in NERVES. Whilst BASICS did not set a minimum level of expertise for surgeons, details on the surgery log were collected with regards to the composition and expertise of the surgical team to support analyses of treatment provider experience if required. *Recommendation LC-2*, trial specific training, is demonstrated at the surgeon level as in TOPS, where surgeons were calibrated and signed off by an expert surgeon prior to participation. None of the trials presented adopted an expertise-based design (*Recommendation LC-3*) although this is likely to be due to concerns around generalisability. For example, all TOPS centres routinely operated on infants that were closer to one timing arm than the other prior to participation in the study. However, adopting an expertise-based design in this trial would limit the generalisability of the trial results. (82, 83) The trial recruited internationally across five countries and therefore it is likely that practice in terms of age at primary surgery will vary. Allowing surgeons to operate on their preferred arm, or the arm in which they have most experience in, would mean that single countries were operating on one arm of the study. Such as design may invariably lead to imbalances in environmental, socio-economic and treatment factors, across the two arms by design. PANasta shows how monitoring of protocol adherence and treatment delivery can be incorporated into trial design (*Recommendation LC-4*). In this trial, operative photographs were collected, for central review, to enable post-surgery quality assessments to be undertaken. None of the trials explored learning as a main or secondary analysis, *Recommendation LC-5*. However,

investigations into changes in surgical outcomes over time was planned for TOPS and presented in a later chapter of this thesis.

These trials also provide real examples of how clustering could be considered following the recommendations, presented in *Box 9* in *Chapter 3*. All trials balanced the randomisation strata, with BASICS balanced on centre through stratification (*Recommendation C-1*) and TOPS balancing on surgeon through minimisation (*Recommendation C-2*). Following balancing the randomisation, which can introduce correlation in outcomes within strata, the analysis was subsequently adjusted, with NERVES stratifying and subsequently adjusting all analysis of primary and secondary outcomes for recruiting centre. Following stratifying randomisation by the same, BASICS did not subsequently adjust for recruiting centre and instead adjusted for age category of recruiting hospital (paediatric or adult) due to its strong prognostic value and overlap with the stratification factor (*Recommendation C-7)*. Many of the trials also considered stricter protocol requirements for treatment delivery, following *Recommendation C-4*. For example, in TOPS where centres were acting outside of their routine practice, surgeons were required to be standardised by an expert surgeon prior to participation. Likewise, a surgical manual was developed for PANasta, this was developed as part of a pilot phase for the trial and detailed steps of the intervention that were mandatory, prohibited and flexible. These steps led to increased monitoring of protocol adherence and treatment delivery (*Recommendation C-5*) with the operative photographs, taken as part of the PANasta trial to ensure ongoing quality and consistency in treatment delivery, providing an example of managing effects by conduct. Many of the trials also explored treatment effects across centres (*Recommendation C-6*), with such investigations being done either through exploratory analysis, as in BASICS, or through formal adjustment as in NERVES.

It is imported to consider some limitations in terms of the generalisability of this summary. First, all example trials are connected to the Liverpool Clinical Trials Unit meaning that they were developed under a single set of standard operating procedures. However, the trial teams

largely comprised different methodologists and therefore do not represent approaches implemented by a single trial team. Second, the examples were also purposely selected to ensure that they cover each of the types of surgical trials types, and therefore may not be representative of other trials within the same trial type or those developed at other Units. The selection also means that the cohort does not include a trial involving entirely new treatments, where the amount of attention given to aspects such as surgical learning would likely be greater. The selection also does not include a single centre study, where elements of clustering may be less of a concern than in more pragmatic multi-centre studies. However, as identified throughout this and subsequent chapters, clustering (in *Chapter 7*) and surgical learning (in *Chapter 8*) may be present even in trials comparing established techniques. It is important to note that the role of this summary is to demonstrate approaches that have been taken to manage learning and clustering in practice, and not to determine how trials should be designed and analysed which is addressed in earlier and later chapters of this thesis.

6.6. Conclusions

The description of trials in this chapter have demonstrated how learning and clustering can be managed through design, conduct and analysis in a variety of different settings, intervention types and across different surgical specialties. Despite the best efforts of trialists, issues around learning and clustering in many aspects are unavoidable and, through early statistical involvement, trials can be designed to reduce the impact of learning and clustering and, when collect and analyse the data to alleviate concerns around either of these effects.

Given that the current guidelines support exploring heterogeneity of treatment effects across treatment providers, there is also a need to understand, within the trial setting, how and when clustering or learning by treating centre or surgeon impacts trial results. The next two chapters explore the impact of clustering on trial results, presenting methods to appropriately adjust for

these, and learning within a trial setting, presenting exploratory analysis approaches to demonstrate these, respectively.

Chapter 7 : Quantifying the impact of clustering on trial conclusions

7.1. Introduction

In *Chapter 1,* clustering was introduced. Approaches for minimising the impact of clustering through trial design and analysis were provided in *Chapter 2* to *Chapter 5* and *Chapter 6* presents real examples of considering their effect. Although good design can minimise the impact of clustering, it is important to consider any remaining impact on trial conclusions.

In individually randomised trials, it is possible that the effectiveness of the intervention, independent of any treatment effect depends on differences in environmental or socio-economic factors. Additionally, clustering may occur due to treatment provider expertise, and this could be at the centre or surgeon level. (84) The statistical measure of clustering between patients treated by the same surgeon or centre is known as the as the ICC and this can vary substantially trial to trial in surgery. (85) In the presence of clustering, the usual statistical methods for analysing trial data may not be appropriate as they often assume independence. (86) Dependence can lead to inflated standard errors, *p*-values, and wider confidence intervals and a reduction in the effective sample size, which ultimately leads to a reduction in the power of the study. (73, 87)

The aim of this chapter is to explore the presence of clustering, by centre and surgeon, in RCT datasets and determine the impact on trial conclusions. A retrospective statistical analysis of outcomes from two individually randomised example trial datasets is undertaken. Next, the impact that adjusting for surgeon, under different levels of clustering and treatment differences, is explored by simulating data.

7.2. Methods

7.2.1. Datasets

This investigation reanalyses example trials: BASICS and TOPS. Both were introduced in *Chapter 6* and represents trial datasets with potential clustering

The BASICS secondary outcome, failure for any cause, is explored in place of the original primary outcome, failure due to infection. In the original trial, 394 patients had a revision for any cause (secondary outcome) compared to 75 patients with infections (primary outcome). Choosing to reanalyse this secondary outcome allows for a greater event rate to support a multilevel analysis approach and the outcome may also be impacted by the surgeon or centre. To further facilitate this analysis and make best use of the statistical methodology available, revision for any cause is analysed as a binary outcome, and not time to event as in the original trial.

The TOPS secondary outcome, occurrence of fistula, is explored in place of the original primary outcome, insufficient velopharyngeal function at age five years. This is because fistula is highlighted within the cleft literature to be associated with surgeon expertise. (88, 89)

7.2.2. Statistical analysis

Statistical principles and presentation

A complete case analysis approach is performed using the intention-to-treat principle. In keeping with approaches used in the original trial, 97.5% confidence intervals are presented for BASICS and 95% confidence intervals presented for TOPS.

The number of infants undergoing surgery with outcome data is presented, as is the number of infants within surgeon and within centre. Binary data are presented by frequencies and

percentages, and continuous data by medians, inter-quartile range and overall range. Outcomes, by centre and surgeon, are presented overall and split by treatment arm using tables and caterpillar plots.

Statistical modelling

The following four methods to analyse individually RCTs with clustering were considered for this investigation (84, 90):

(1) Cluster level analysis;

(2) Random effects, or cluster specific approach;

(3) Marginal or population-averaged approach; and

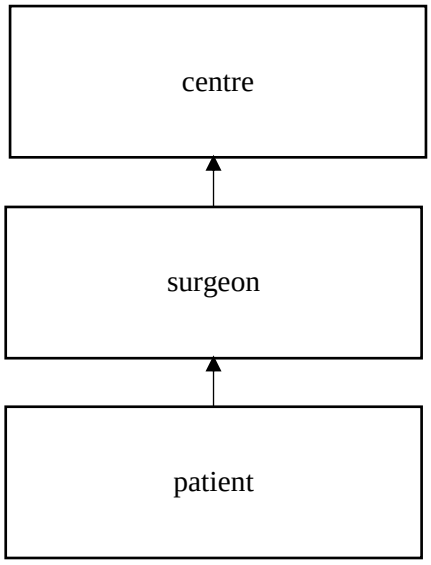
(4) Ignoring the effect and assume independent outcomes.

(1) does not make optimal use of the data and does not allow adjustment of individual patient factors. (2) is preferable to (3) as it allows the effect of a treatment provider to be explored on an individual level and estimates the ICC, which enables heterogeneity to be more fully explored. (4) represents the common approach used within trials. This analysis therefore compares and contrasts methods (2) and (4).

The outcomes for both trials are binary and the data are multilevel, with patient clusters nested within surgeon nested within centre. *Figure 5* presents a classification diagram of the three-level data hierarchy.

Four models are be applied to each dataset, all models adjust for treatment and have the response as the outcome. *Model A* represents method (4) in which the effects of clustering are ignored. The second, third and fourth models, *Model B, Model C* and *Model D* respectively, are random effects models as per method (2), each with varying complexity accounting for the different hierarchies in the data due to centre and surgeon.

Figure 5: Classification diagram for the three-level example



Model A is a simple logistic regression model, modelling the chance of an event based on the treatment.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i$$

Equation 1

Where π indicates the probability of the i-th patient having the event, β_0 is intercept, and β_1 is the regression coefficient for treatment.

Model B is a two-level multilevel logistic model. This model analyses the two-level data structure of patients (level 1) within centre (level 2).

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + u_j \quad \text{where } u_j \sim N(0, \sigma_u^2)$$

Equation 2

As in *Equation 1*, β_1 regression coefficient for treatment is now the effect of x after adjusting for centre effect u. While β_0 is the overall intercept, the intercept for a given centre *j* is $\beta_0 + u_j$, where u_j can be referred to as the group (random) effect, group residual or level 2 residual.

Model C is also a two-level multilevel model. This model analyses the two-level data structure of patients (level 1) within surgeon (level 2). This model is the same as the centre adjusted two-level multilevel logistic model in *Equation 2*, with surgeon as the group effect in place of centre.

Model D is a three-level hierarchical model. This analyses the three-level data structure of patient (level 1) within surgeon (level 2) within centre (level 3) in its entirety. This model has been included, as naively fitting two-level models to three-level data can lead to misattributing response variation to the two included levels. (91)

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \beta_0 + \beta_1 x_{ijk} + u_{j(k)} + u_k$$

Equation 3

As in *Equation 1*, β_1 regression coefficient for treatment is now the effect of x after adjusting for centre effect u_k and surgeon within centre effect $u_{j(k)}$. While β_0 is again the overall intercept, β_1 denotes the regression parameters for the patient level variable treatment. $u_{j(k)}$ denotes the random effect for the *j*-th surgeon within the *k*-th centre and u_k denotes the random effect for the *k*-th centre. Assuming that $u_{j(k)} \sim N(0, \sigma^2_{u_1})$ and $u_k \sim N(0, \sigma^2_{u_2})$ are independent.

Estimating the ICC

In a two-level multilevel linear regression model of a continuous outcome the intra-level-2 correlation is:

$$\hat{\rho} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

Equation 4

Where σ_u^2 and σ_e^2 denotes the between-cluster variation and the between-patient variation respectively. When applying variance components models, this also measures the proportion of the total variances which is between level 2 units, such as surgeon. In a three-level multilevel model, patients within surgeons within centres, there are two such correlations and variance proportions; the intra-centre correlation which is also the proportion of variance that is between centres and the intra-surgeon correlation, which is also that between surgeons.

The proportion of variance explained by both centre and surgeon is estimated using *Equation 2*. The between-cluster variance (σ_u^2) is estimated automatically when applying the SAS PROC GLIMMIX procedure. (92) However, when using logistic regression models, the between-subject variance (σ_e^2) cannot be directly computed. A number of procedures for calculating the ICC within this modelling approach exist. The most common being the latent response formulation which is inappropriate for this data as assumes that the underlying distribution of the outcome is continuous. While there is debate around application within binary outcomes, for the outcomes considered within this investigation the argument for a continuum between the states of presence and absence is difficult to apply. (93) Instead, the between-subject variance is estimated using simulations within SAS. (Statistical Analysis Software (SAS®) 9.1.4; SAS Institute Inc., Cary, NC, USA) (93) (*Appendix Material 6*) The simulations method does not require an approximating formula. Using the intercept and σ_u^2 parameters obtained from *Equation 2*, and simulating 50 000 values for the level 2 residual, the coefficient σ_e^2 is estimated. These are then used to calculate $\hat{\rho}$ as in *Equation 4*. (94)

Simulations

The performance of the adjusted and unadjusted models when applied to the individually randomised trial data is explored using Monte Carlo simulation methods, with the aim of exploring the impact that adjusting has on the null hypothesis under varying degrees of ICC and treatment differences.

Data are simulated using the TOPS dataset, as this dataset contained fewer surgeon clusters and each cluster is of reasonable size thereby reducing the chance of the models failing to converge. All derived datasets have the same sample size, number of surgeons and patients within surgeon as observed in the TOPS trial. Scenarios allowing for varying ICC and odds ratio are explored. To simulate surgeon effects, the relationship between ICC and between-patient variation (σ_e^2) in *Equation 4* is employed, with the between-cluster variation (σ_u^2) varying and the between-patient variation set to that observed in TOPS. The following ICC levels are considered for completeness: 0.001, 0,01, 0.05, 0.10, 0.15, 0.20 and 0.25 alongside the ICC level observed within the original trial. The choice of ICC variations is weighted towards zero as lower values occur in practice. The following odds ratios are considered: 0.6842, 0.4004, 0.2416, to represent low, medium and large effect sizes in samples where the rate of the outcome is approximately 10%, alongside the odds ratio observed within the original trial. (95) Data are simulated for each scenario using *Model C.* Treatment allocations for patients within surgeon are assigned using simple randomisation with an equal allocation ratio, using $X_j \sim Bernouilli(0.5)$. Outcomes for each patient are generated using the logit link function applied to a linear predictor, utilising a random effect for surgeon based on the ICC for that scenario. (96)

Model A and *Model C* are applied to each simulated scenario. The TOPS randomisation was adjusted for surgeon and so *Model* C is a logical choice over *Model B.* Most centres had a single surgeon and due to this overlap, *Model D* was dropped for the simulation aspect of this

study. Also, as the randomisation schedule used surgeon, this clustering is non-ignorable by design.

2000 simulated datasets is generated for each scenario. The two models, *A* and *C*, are applied to each and their estimated parameters extracted and used to estimate empirical power and empirical coverage. These measures target the null hypothesis directly, as the odds ratios cannot be compared directly, due to *Model A* providing a marginal and *Model C* a conditional estimand. Empirical power is defined as the proportion of simulations per scenario in which the estimated 95% confidence intervals of the estimated odds ratio does not include 1 (no difference). Empirical coverage of confidence intervals is defined as the proportion of simulations per scenario in which the estimated 95% confidence interval of the estimated odds ratio includes the true odds ratio. All datasets are simulated and analysed in SAS. (Statistical Analysis Software (SAS®) 9.1.4; SAS Institute Inc., Cary, NC, USA)

7.3. Results

7.3.1. BASICS

Table 22 provides a summary of the dataset. *Table 23* provides the model parameters used to estimate the ICC presented in *Table 22*. *Table 24* presents the number of units at each level of the data hierarchy. The number of centre clusters was 21, the number of surgeon clusters was 303. Surgeons operated up to a maximum of three centres, yet generally cross centre operations were infrequent, see *Table 22*.

The within centre and surgeon cluster size varied in BASICS, with 21 centre clusters containing 303 surgeon clusters. The median cluster size for patients within centre and surgeon was 71 and 3 respectively. Almost all centres operated on more than ten patients (n=19/21), compared to 13.5% of surgeons (n=41/303). *Table 22* and *Table 24* provides more details.

Table 22: Summary of examples of trials with the potential for clustering

	BASICS	TOPS
Sample size	1594	552
	Standard : 533 (33.4%)	6 months : 279 (50.5%)
	Antibiotic : 535 (33.6%)	12 months : 273 (49.5%)
	Silver : 526 (33.0%)	
No. patients with event	398 (25.0%)	73 (13.2%)
	Standard : 130 (24.4%)	6 months : 40 (14.3%)
	Antibiotic : 132 (24.7%)	12 months : 33 (12.1%)
	Silver : 136 (25.9%)	
No. of centre clusters	21	22
Median cluster size	71.0	17.0
No. at least 10 patients	19	15
ICC estimate	Standard : 0.0840	6 months : 0.0152
	Antibiotic : 0.0843	12 months : 0.0133
	Silver : 0.0862	
No. of surgeon clusters	303	26
Median cluster size	2.0	15.5
No. at least 10 patients	41	17
ICC estimate	Standard : 0.0840	6 months : 0.0217
	Antibiotic : 0.0826	12 months : 0.0190
	Silver : 0.0840	
No. of centres per surgeon		
One	271 (89.4%)	20 (76.9%)
Two	31 (10.2%)	6 (23.1%)
Three	1 (0.3%)	0 (0.0%)

Table 23: Parameter estimates used to estimate the ICC

Reference level for treatment: BASICS – standard; TOPS - twelve-months surgery.

	BASICS	TOPS
Centre model		
Intercept	-1.0672	-1.9343
Treatment	Antibiotic : 0.01411	6 months : 0.1952
	Silver : 0.08843	
σ_u^2	0.4982	0.1201
Surgeon model		
Intercept	-1.0227	-1.9103
	Antibiotic : -0.03065	6 months : 0.2069
	Silver : 0.02267	
σ_u^2	0.4870	0.1676

Table 24: Number of units at each level of the data hierarchy

Level number	Level	BASICS	TOPS
3	Centre	21	22
2	Surgeon	303	26
1	Patient	1594	552

The ICC statistic for patients on the standard arm is estimated to be 0.0840 and 0.0840 for centre and surgeon respectively, see *Table 22*. The ICCs are very similar for the antibiotic and silver arm. This suggests some clustering in the data at these levels, with 8% of the variation in the propensity to require a revision lying between centre or surgeon. *Figure 6* and *Figure 7*, which presents the centre and surgeon level event rates, shows variation further suggesting that there may be some clustering at these levels. The revision rate within each centre varies from 4.8% (*Centre 20*) to 75.0% (*Centre 18*), see *Figure 6*. Likewise, the revision rate within

each surgeon varied, see *Figure 7.* Shunt failures overall in BASICS, and for each shunt type, by centre and surgeon is presented in *Appendix Table 6* and *Appendix Table 7* respectively.

Figure 6: Caterpillar plot of failure rates by centre in BASICS

Within centre revision rates. 97.5% confidence intervals are plotted and compared against the overall failure rate of 25.0%.

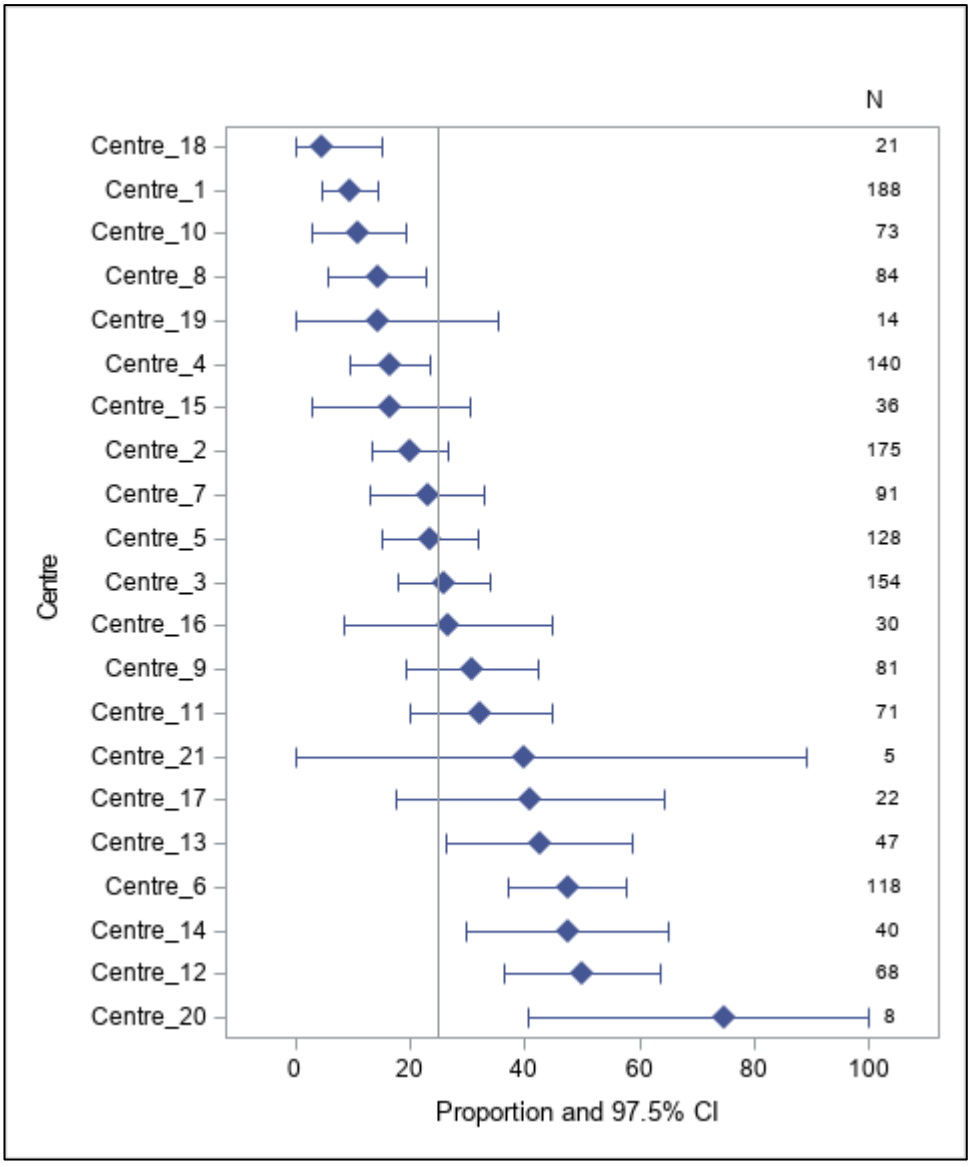


Figure 7: Caterpillar plot of failure rates by surgeon in BASICS

Within surgeon revision rates. Includes surgeons who operated on at least ten patients only. 97.5% confidence intervals are plotted and compared against the overall failure rate of 25.0%.

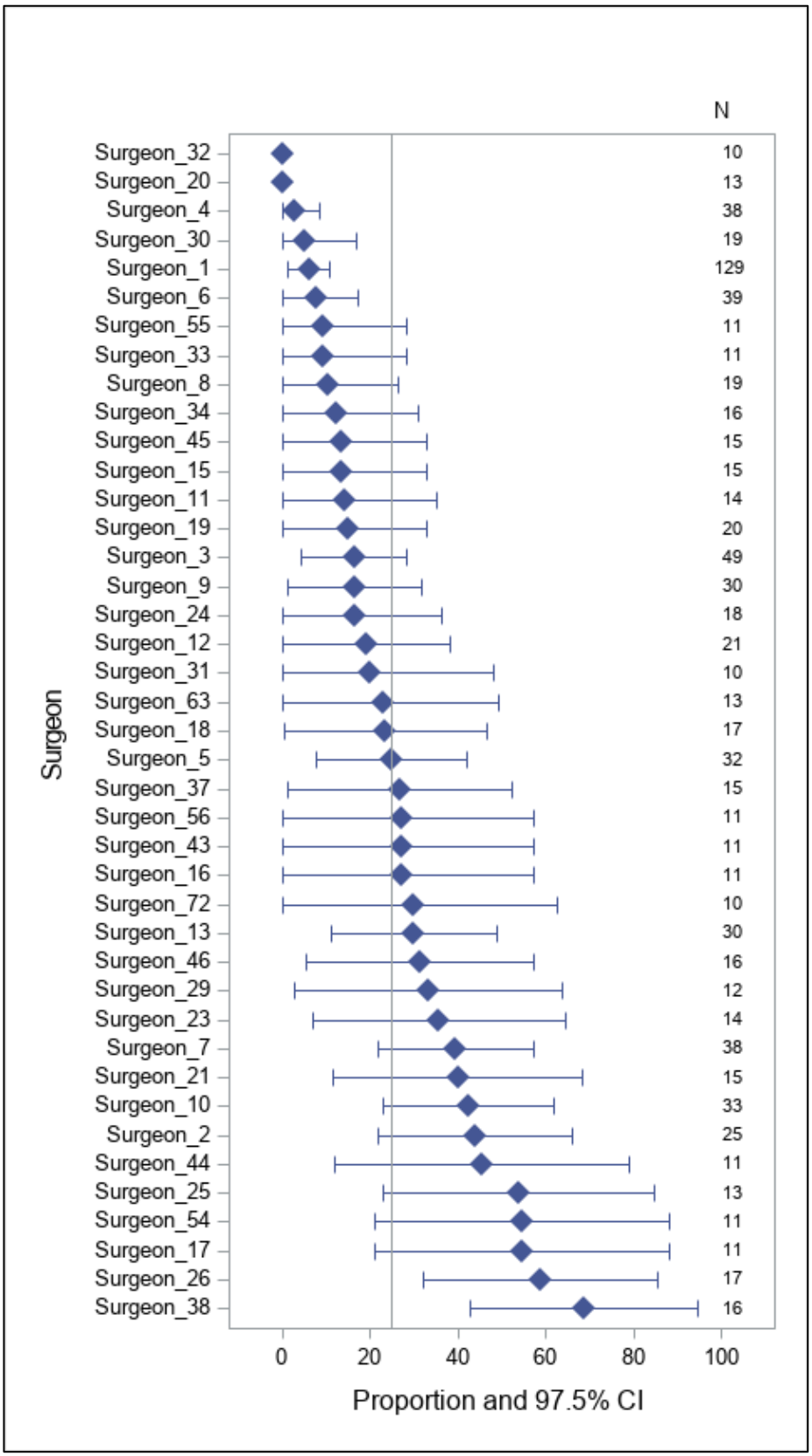


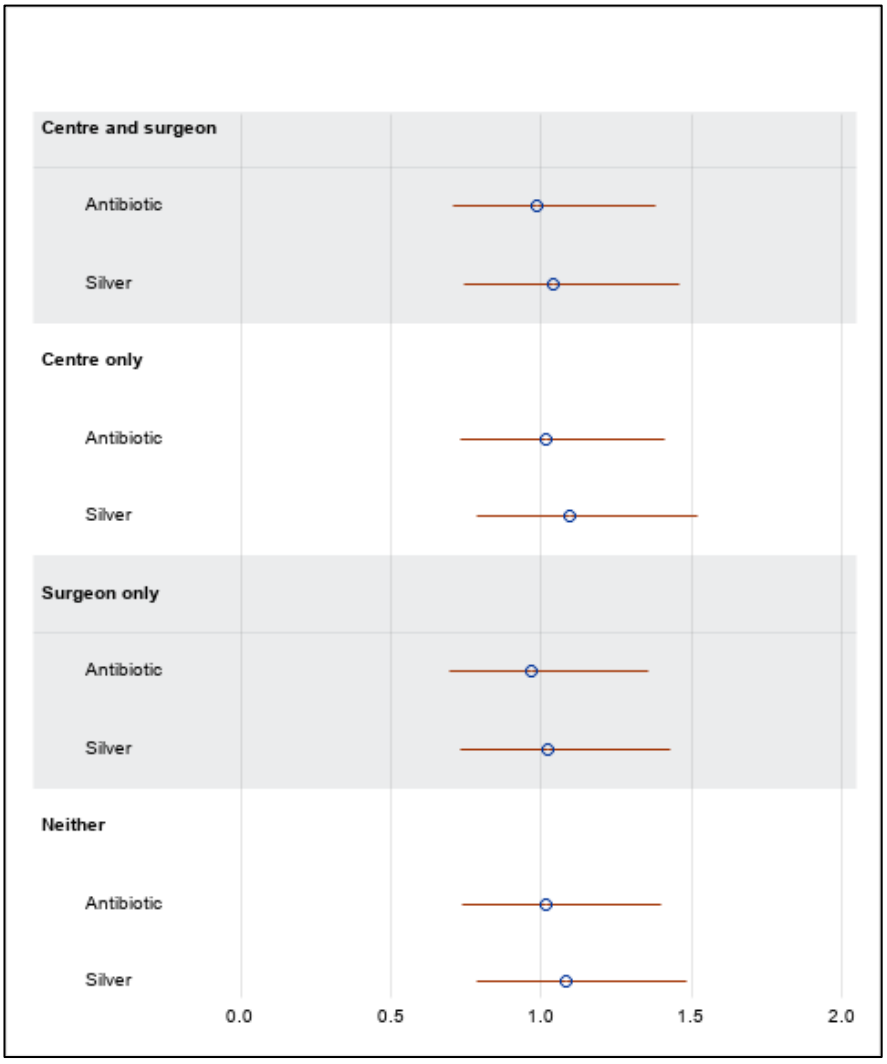
Table 25 and *Figure 8* shows the results of the model fitting for each adjustment approach. The conclusions made by all four approaches is that there is no difference between the antibiotic impregnated and silver shunt, when compared to the standard shunt. With the surgeon adjusted model giving almost identical odds ratios for both treatment effects, compared to the other three adjustments where silver appeared to be marginally inferior to standard. The size of the confidence intervals varied very little across each approach.

Table 25: Results of model fitting

Trial	Cluster adjustment (Model)	Odds ratio	Confidence interval [A]	*p*-value
BASICS	*Reference = Standard*			
	Centre and surgeon (D)			
	Antibiotic	0.986	0.704 to 1.381	0.9267
	Silver	1.043	0.745 to 1.461	0.7788
	Centre only (B)			
	Antibiotic	1.014	0.729 to 1.411	0.9236
	Silver	1.092	0.786 to 1.518	0.5467
	Surgeon only (C)			
	Antibiotic	0.970	0.694 to 1.356	0.8374
	Silver	1.023	0.732 to 1.430	0.8794
	Neither (A)			
	Antibiotic	1.015	0.738 to 1.397	0.9145
	Silver	1.081	0.787 to 1.486	0.5826
TOPS	*Reference = 12 months*			
	Centre and surgeon (D)	1.232	0.744 to 2.041	0.4167
	Centre only (B)	1.216	0.740 to 1.998	0.4405
	Surgeon only (C)	1.230	0.743 to 2.036	0.4203
	Neither (A)	1.217	0.742 to 1.998	0.4363

[A] 97.5% presented for BASICS; 95% presented for TOPS.

Figure 8: Results of model fitting for BASICS



7.3.2. TOPS

Table 22 provides a summary of the dataset. *Table 23* provides the model parameters used to estimate the ICC presented in *Table 22. Table 24* presents the number of units at each level of the data hierarchy. The number of centre clusters was 22, the number of surgeon clusters is 26. There was little overlap of surgeons across centres, with six surgeons operating across different centres, see *Table 22*. The number of cases outside of their primary centre were very few.

Overlap was evident between the within centre and surgeon clusters in TOPS, with 22 centre clusters containing 26 surgeon clusters. The median cluster size for patients within centre and surgeon was 17 and 15.5 respectively. Approximately two-thirds of all centres (n=15/22) and surgeons (n=17/26) operated on more than ten patients. *Table 22* and *Table 24* provides more details.

The ICC statistic for infants in the six months arm is estimated to be 0.0152 and 0.0217 for centre and surgeon respectively, see *Table 22*. The ICCs are very similar for the twelve months arm. This number suggests that there is little clustering. However, *Figure 9* and *Figure 10* shows that the fistula rate does differ at the centre and surgeon level respectively. The fistula rate within each centre varies from as no cases in five centres, to 50.0% in *Centre 19*, where caseload was low, see *Figure 9*. Likewise, the fistula rate within each surgeon varied, see *Figure 10*. However, the sample was small in some cases. Overall fistula rates in TOPS, and for each surgery timing, by centre and surgeon are presented in *Appendix Table 8* and *Appendix Table 9* respectively.

Table 25 and *Figure 11* shows the results of the model fitting when using the various adjustment approaches. For all four approaches, the treatment effect is not statistically significant at the 5% level. The centre and surgeon adjusted model and surgeon only model treatment estimates are very similar, as are the centre only model and the model with neither adjustment. Whilst the results suggest no difference in fistula rates between the groups, the six months arm patients appear to have higher odds of a fistula when compared to the twelve months patients. As with the BASICS analysis, the confidence intervals around the treatment effect are very similar across each approach.

Figure 9: Caterpillar plot of fistula rates by centre in TOPS

Within centre revision rates. 95% confidence intervals are plotted and compared against the overall failure rate of 13.2%.

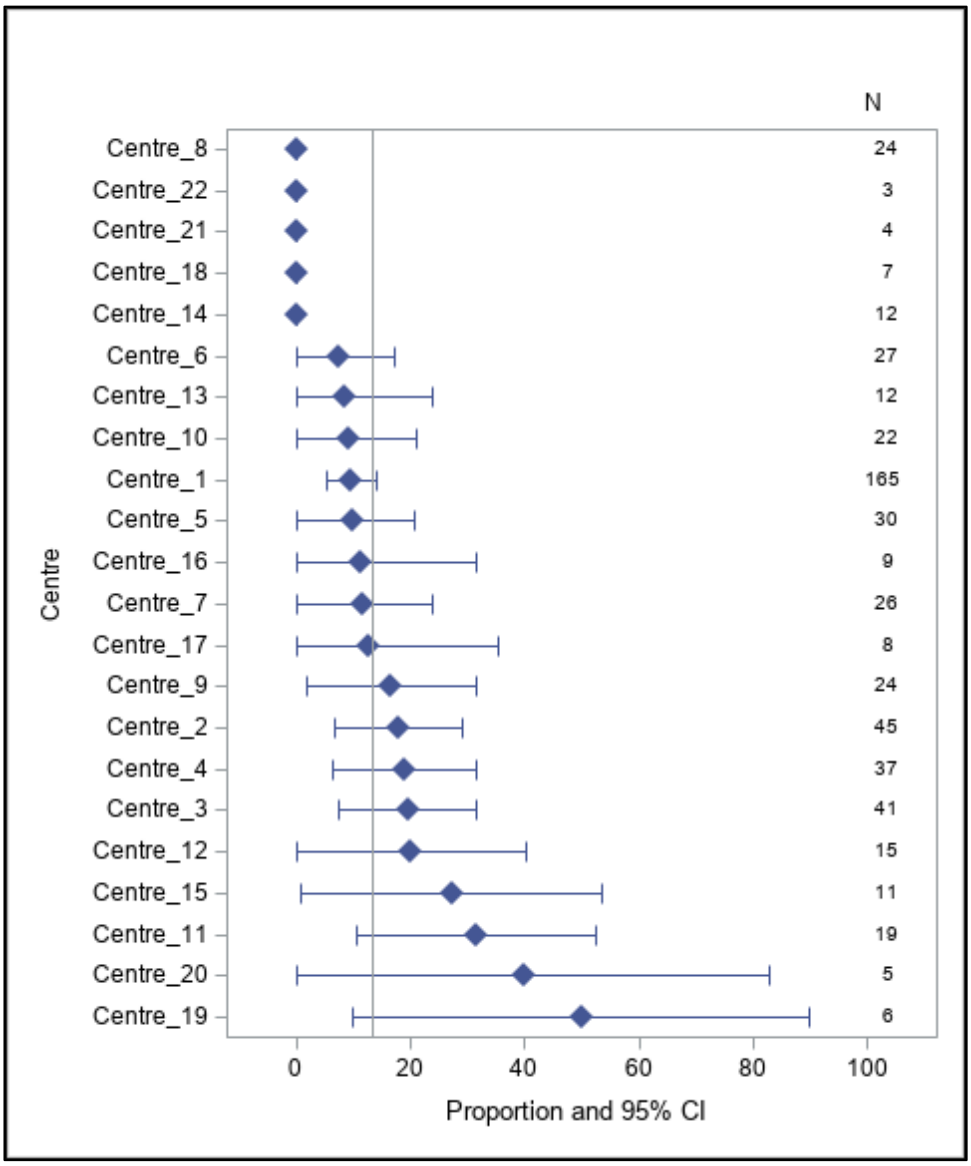


Figure 10: Caterpillar plot of fistula rates by surgeon in TOPS

Within surgeon revision rates. 95% confidence intervals are plotted and compared against the overall failure rate of 13.2%.

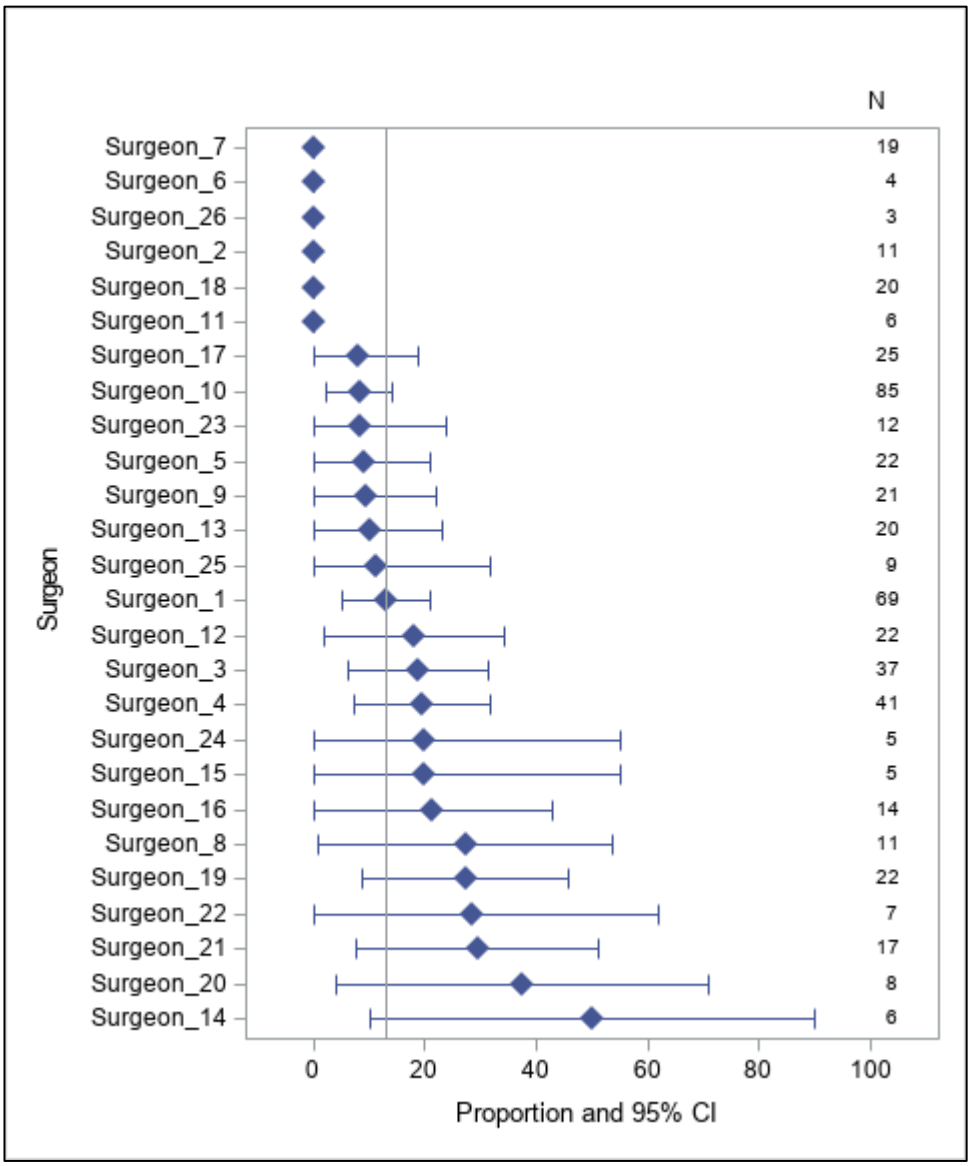
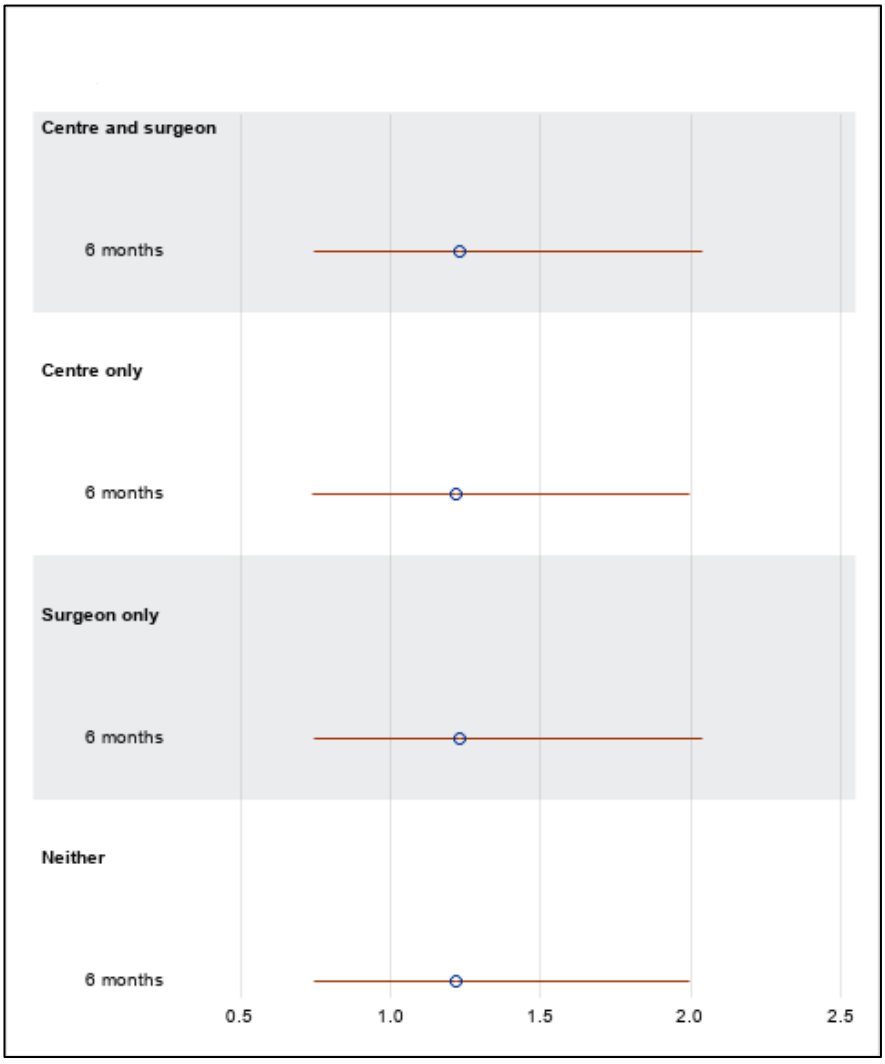


Figure 11: Results of model fitting for TOPS



7.3.3. Simulations

2000 datasets were generated for each of the 32 scenarios, see *Appendix Material 7*. The fixed parameters were those estimated by applying *Model C* to the original TOPS data, presented in *Table 23* and *Table 25*. Each of the 32 scenarios had *Model A* and *Model C* applied, see *Appendix Material 8* and *Appendix Material 9*, and from this the empirical power and empirical coverage derived.

The empirical power was very similar, less than 0.6% difference, for the adjusted and unadjusted models for lower values of the ICC. As the ICC increased, the adjusted model

provided better power, up to 5%, particularly for small or medium odds ratios. *Table 26* and *Figure 12* provides further detail.

Table 26: Empirical power of all scenarios

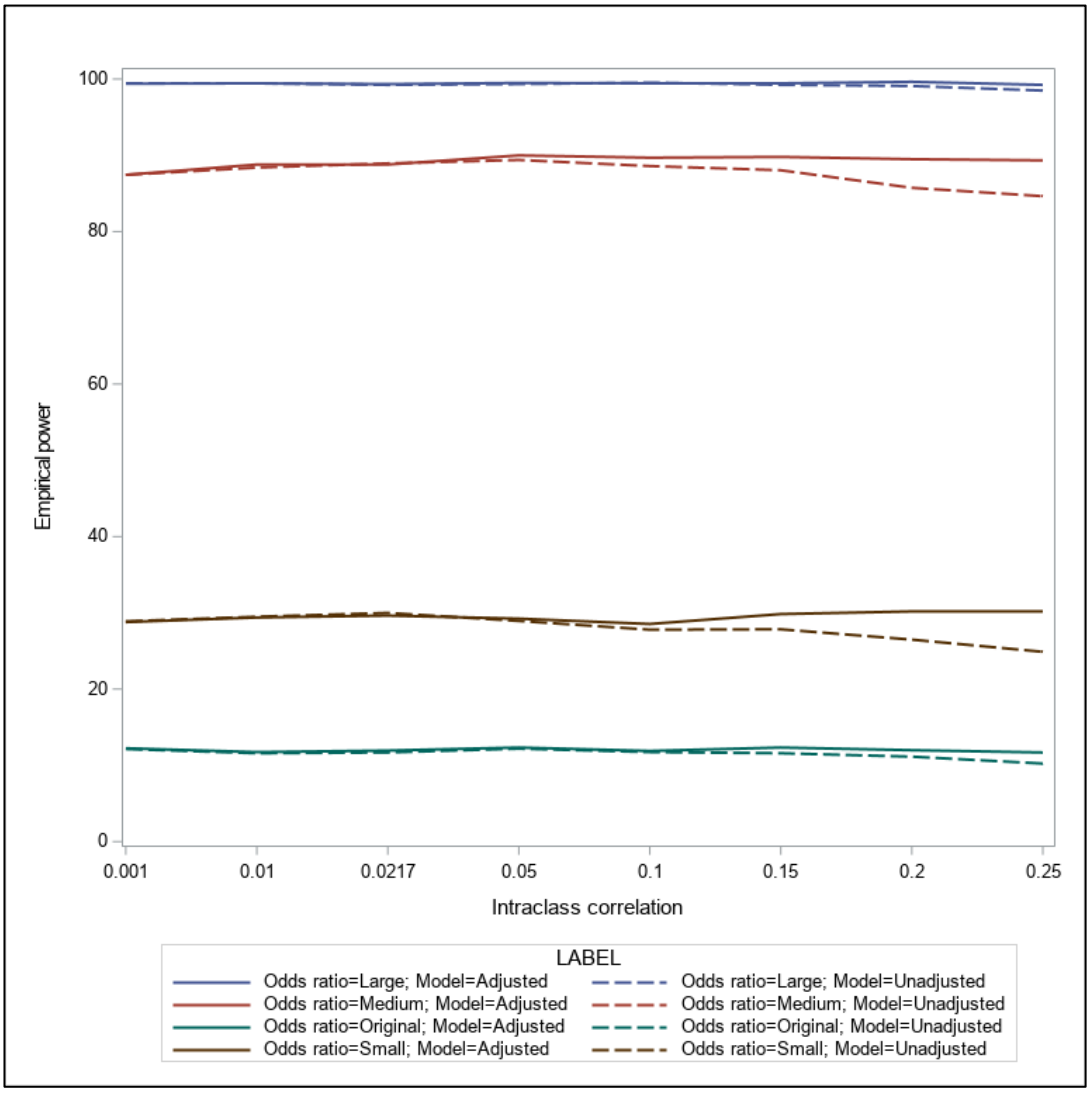
Odds ratio [A]	Model [B]	Intraclass Correlation Coefficient							
		0.001	0.01	Original [C]	0.05	0.10	0.15	0.20	0.25
Original	Adjusted	12.25	11.75	11.95	12.35	11.90	12.35	12.00	11.70
	Unadjusted	12.15	11.60	11.70	12.20	11.75	11.60	11.15	10.25
Small	Adjusted	28.80	29.40	29.65	29.25	28.55	29.85	30.20	30.20
	Unadjusted	28.90	29.50	30.00	28.95	27.80	27.85	26.50	24.90
Medium	Adjusted	87.45	88.80	88.80	90.00	89.70	89.80	89.50	89.35
	Unadjusted	87.45	88.40	88.95	89.40	88.60	88.05	85.75	84.65
Large	Adjusted	99.40	99.45	99.35	99.50	99.45	99.45	99.65	99.25
	Unadjusted	99.45	99.45	99.25	99.35	99.55	99.25	99.10	98.50

[A] **Odds ratios -** Original: 0.813100956; Small: 0.6842285323; Medium: 0.4004485023; Large: 0.2416217653.

[B] **Model –** Adjusted: Model C - Multilevel logistic regression model with a random effect for surgeon; Unadjusted: Model A - Simple logistic regression model.

[C] Original ICC: 0.0217

Figure 12: Empirical power of all scenarios



The empirical coverage was very similar, less than 0.5% difference, for the adjusted and unadjusted models when the ICC was less than 0.05. As the ICC increased beyond 0.1, the adjusted model provided better coverage for all odds ratios and peaked for the largest ICC and odds ratio where 94.4% of unadjusted models contained the true odds ratio compared to 75.65% of the unadjusted models. *Table 27* and *Figure 13* provides further detail.

Table 27: Empirical coverage of confidence intervals for all scenarios

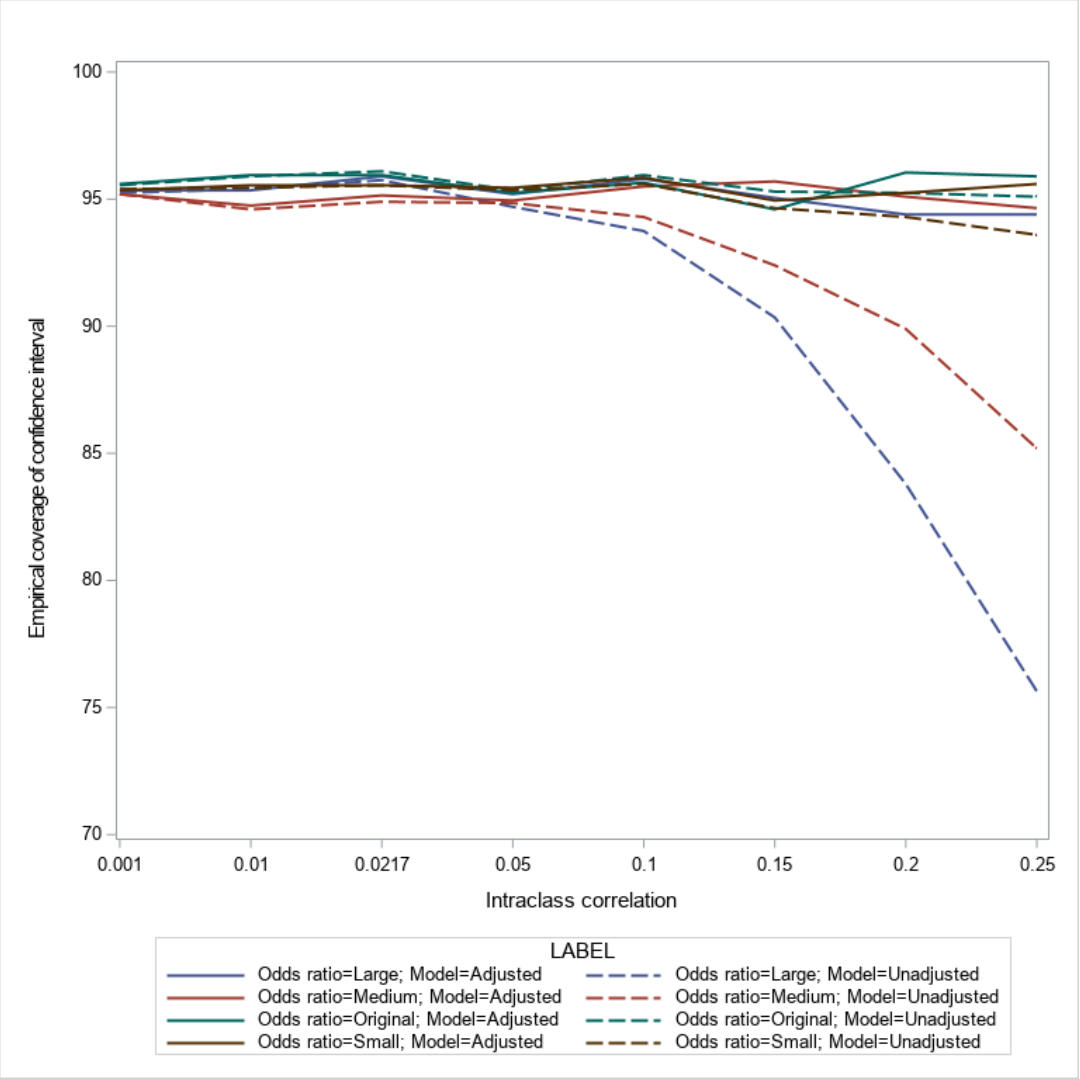
Odds ratio [A]	Model [B]	Intraclass Correlation Coefficient							
		0.001	0.01	Original [C]	0.05	0.10	0.15	0.20	0.25
Original	Adjusted	95.60	95.95	95.95	95.25	95.65	94.60	96.05	95.90
	Unadjusted	95.55	95.90	96.10	95.35	95.95	95.30	95.25	95.10
Small	Adjusted	95.35	95.55	95.55	95.45	95.85	94.95	95.25	95.60
	Unadjusted	95.40	95.45	95.55	95.35	95.60	94.65	94.30	93.60
Medium	Adjusted	95.20	94.75	95.15	94.95	95.50	95.70	95.10	94.65
	Unadjusted	95.20	94.60	94.90	94.85	94.30	92.40	89.90	85.20
Large	Adjusted	95.35	95.35	95.90	95.20	95.80	95.05	94.40	94.40
	Unadjusted	95.25	95.40	95.75	94.70	93.75	90.35	83.80	75.65

[A] **Odds ratios -** Original: 0.813100956; Small: 0.6842285323; Medium: 0.4004485023; Large: 0.2416217653.

[B] **Model –** Adjusted: Model C - Multilevel logistic regression model with a random effect for surgeon; Unadjusted: Model A - Simple logistic regression model.

[C] Original ICC: 0.0217

Figure 13: Empirical coverage of confidence intervals for all scenarios



7.4. Discussion

The estimated treatment effect, and associated confidence intervals, of the four models were similar. The conclusions reached using the adjusted models matched those made with the unadjusted models. This is supported by a simulation study where smaller ICC values and smaller odds ratios were associated with smaller differences between adjusted and unadjusted models.

Both datasets had a moderate sample sizes (BASICS: 1594; TOPS: 552) and small ICCs at both the centre and surgeon level (BASICS < 0.09; TOPS < 0.02). There was much more

overlap between centre and surgeon clusters in TOPS than BASICS. The within surgeon cluster size was more varied in BASICS than TOPS, with the median cluster size being smaller in BASICS (BASICS: n=3, TOPS: n=15.5) and a larger number of surgeons operating on more than ten patients (BASICS: 41, TOPS: 17).

The within centre and surgeon sample size when there are a large number of clusters compared to the overall sample size should be considered when undertaking analyses that adjust for cluster effects and this was the basis for simulating data using the TOPS dataset instead of the BASICS dataset. Large numbers of clusters compared to the sample size can lead to biased estimates or inflated type I error rates, particularly with binary outcomes where there are too few events per cluster. (97)

This analysis did not explore the cross membership of surgeons in centres. However, BASICS recruited over four years and TOPS over seven, the nature of the medical community often leads to surgeons moving or working between centres leading to cross membership between clusters i.e. surgeons not unique to a centre. Whilst there were some cross membership in these datasets, surgeon had a majority centre. The number of operations at other centres were largely single cases, and therefore minimal compared to the total number of operations undertaken as part of the study. Whilst cross membership was not explored as part of this analysis, it is summarised for transparency. The analysis undertaken makes use of the available methodology at this time and methods to allow for such explorations of binary outcomes across three cluster levels should be considered as an area for future work.

Another potential issue is related to variation in sample size between clusters at both the surgeon and centre level. A point of view which is often propounded is that number of patients per centre be roughly equal, and in surgical trials the role of the centre can also be extended to the surgeon. Reasons for this view include inefficiency in terms of the precision of the estimate delivered at the end of the trial. The impact on treatment effect is not explored here,

this is to ensure that the methods applied are generalisable. It is unlikely that real world trials have equal group sizes. Instead, the simulations were designed to represent a study with varying cluster sizes to mimic real world application.

Treatment by centre, or by surgeon, interaction terms were not considered when undertaking this analysis. First, this was due to the efforts in designing these studies, summarised in *Chapter 6*, to ensure standardisation of the treatments delivered in terms of defining uniform centre, surgeon and patient selection criteria. Second, the issues around adequate sample size within clusters is highlighted above. Whilst sample size is a concern when making-adjustments at the levels explored here, a larger sample than that within these studies would be required to undertake treatment by cluster analyses and such sample sizes are rare in RCTs. In fact, as discussed in *Chapter 2*, the use of interaction terms should be avoided for primary analyses in these settings and an exploratory analysis of changes in treatment by deliverer is instead explored in *Chapter 8*.

For this analysis, all operations have been analysed according to a single principal surgeon. Particularly within BASICS, the number of surgeons present varied, with each additional operator introducing an additional potential source of variation that may require management. The multiple members of the operating team, the potential impact of the anaesthetist, and the differing level of involvement (scrubbing) and expertise (grading) may impact results and has not been considered here for reasons around sample size and generalisability for methods to be applied to other RCT data. However, as a further exploration, the impact of the wider team may require consideration with regards to the assumptions and weighting applied prior to undertaking modelling.

Ignoring clustering effects leads to inflated standard errors and loss of statistical power. (98) This reanalysis of trial data demonstrates little difference in the estimated treatment effects when adjusting for centre and surgeon effects compared to taking an unadjusted approach,

although this only applies to studies observing smaller ICCs and treatment differences, as supported by the simulation study. The need to account for clustering becomes vital as the odds ratio or ICC increases, where the true difference can be missed by unadjusted analysis methods.

7.5. Conclusions

Despite the similarity of results for our datasets, researchers should be aware of the potential for clustering by centre, or surgeon, when undertaking RCTs and drawing inference from the data collected. Awareness of the potential for clustering will improve data collection methods, which in turn will allow for adjustments to be made or investigated in trials where such adjustments could change the conclusions made.

Chapter 8 : Investigating the presence of surgical learning within a trial setting

8.1. Introduction

In *Chapter 1* learning curves were introduced. Approaches for minimising the impact of them through trial design and analysis were provided in *Chapter 2* to *Chapter 5* and *Chapter 6* presents real examples of considering this effect. As with clustering, good design can minimise the impact of the learning curve it is important to consider any remaining impact on trial conclusions.

In every day practice, it is likely that surgeons will use techniques with which they are most proficient or familiar. Learning can continue over a very long time for some techniques, perhaps hundreds of procedures and attaining the experience required prior to running a trial may not be possible in some specialties, such as when the condition is rare. (14, 89) Learning can compromise the validity of the trial if the expertise of the surgeon is skewed toward the better established, more widely used, or easy to perform technique. (20, 26, 99)

A statistical description of any learning curve effect is the best way to investigate learning within a trial. (17) When conducting such an investigation, choosing a measure of learning can be difficult. There are generally two types used: measures of patient outcome, such as complication rate, and measures of surgical process, such as length of operation. (17)

This chapter presents an investigation into the surgical learning curve within a clinical trial setting. The aim is to demonstrate how statistical methods can be used to explore for the presence of surgical learning and how, if necessary, their presence can be controlled for within the trial analysis.

8.2. Methods

8.2.1. Learning within the TOPS trial

The TOPS trial was first introduced in *Chapter 6.* At the time of writing, the cleft surgical learning curve has not yet been investigated. Several theories exist around surgeries delivered by high volume surgeons achieving better outcomes, and that individual skill and protocol complexity are also important contributors to surgery success. (88, 100) Defining a minimum number of operations as a pre-requisite for participation in a trial, at levels defined in other fields of surgery, was not possible in TOPS. This is because the amount of experience required, using other specialties as a basis for estimating, would be unrealistic for many surgeons working in the developed world where cleft incidence is lower. (89) To standardise the surgical skills and surgery performed within the TOPS trial through design, participating surgeons underwent training prior to participation. (74)

The TOPS trial aimed to determine whether better speech outcomes could be obtained by conducting the primary surgery to repair the cleft at six-months or twelve-months of age. The majority of the 22 participating centres had one consultant surgeon delivering the trial surgeries and a single method for cleft palate closure was selected, the Sommerlad technique, which provided some standardisation regarding experience. Not all surgeons had previously used this method, and technique training and calibration were provided for all surgeons prior to participation providing further standardisation.

Within the TOPS trial, as well as a learning curve related to the surgical method, learning may also be present within each treatment arm due to differences in time point of delivery. These differences mean that the size of the infant at the time of the operation different between the randomisation arms. Therefore, surgeons who are more experienced in performing the procedure on infants at age six-months may find surgery on infants at twelve-months of age

more difficult, or vice versa. It may also be argued that, surgeons who routinely operated on infants at nine months of age may find the transition to alternative timings less of a challenge.

8.2.2. Statistical analysis

Variables to represent experience

Three variables to represent prior existing, or changing, experiences are coded for the analysis:

1) Experience in cleft palate surgery;
2) Experience in the use of the Sommerlad technique;
3) Experience in operating on infants age six or twelve-months.

1) Experience in cleft palate surgery (X1)

Data on number of procedures performed prior to and outside the trial were not available, however surgical experience within the trial was considered by coding the order in which infants had been treated within the trial (*operation sequence*). This is the closest proxy to a time varying learning effect when no information on non-randomised cases is available and mirrors approaches used in other explorations of learning in randomised trial data. (17, 20)

2) Experience in use of the Sommerlad technique (X2)

Technique experience, that is whether or not the surgeon had experience in delivering the Sommerlad technique prior to participation, is coded as a binary variable.

3) Experience in operating on infants age 6 or 12 months (X3)

Experience relating to the age of the infant (*age experience*) was not available for each surgeon. However, a proxy for this is the age at which primary surgery is given, prior to the trial, at the centre in which the surgeon is based. Where primary surgery is delivered across a range of ages, the median age is used.

Outcomes

Two outcomes highlighted within the cleft literature to be of interest with respect to surgical learning were selected: (88, 89)

1) Operation time (Y1): an outcome of surgical process. Operation time is calculated as the difference in minutes between operating start time (knife to mucosa) and end time (mucosa closure).
2) Occurrence of fistula (Y2): used as a patient outcome of surgery reflecting surgical success. Fistula is defined as in the original trial as a dichotomous outcome of whether or not the child had a postoperative fistula up to five year follow up. (76) [*Conroy, 2020*]

Statistical principles and visual methods

A complete case analysis approach is used following the intention-to-treat principle using a complete case analysis set. (101) All randomised infants, with complete surgery data, were analysed according to the group they were originally randomised to. No imputation methods for missing data were applied. 95% confidence intervals are presented throughout.

The number of infants randomised, undergoing surgery, and completing follow up for outcome data are presented. The hierarchical nature of the data, such as the number of infants within surgeon and within centre is also provided.

Continuous data are presented as means, standard deviations and overall range to be consistent with available methodology to explore learning over time. In the presence of skewed data, a log transformation is considered. Operation time, as a continuous outcome, is presented within surgeon and split by treatment arm. Operation time, split by surgeon, is presented using funnel plots, assigning a target operation time as the overall surgeon mean observed and two prediction limits, at 95% and 99.8%, to identify outlying areas of differing extremity. Prediction limits are confidence intervals around the target operation time, calculated using

the appropriate z-scores for the line represented and each possible *N*. (102) Moving averages, of order five, for surgeons performing at least twenty primary operations are presented. (20)

Binary data are presented as frequencies and percentages. Fistula, as a binary outcome, is presented within surgeon and split by treatment arm. Fistula, split by surgeon, is presented using caterpillar and funnel plots. (102) Trends in experience are investigated using the cumulative sum (cusum) procedure, a graphical method for identifying trends in data, a cusum chart is presented for surgeons performing at least twenty primary operations following the methods presented by *Ramsey et al.*. (20) In the plots, the x-axis represents the operation sequence and the y-axis is a performance indicator based on the within surgeon data series. For a series of observations over time, $\{(X_1, \ldots, X_i): i = 1, 2, \ldots, n\}$, where X_i are dichotomous, then the cusum series is defined as:

$$s_i = \sum_{j=1}^{i} (X_j - X_0)$$

Equation 5

Where X_0 is the predetermined reference levels, representing the desirable performance levels based on the cleft literature, of 85% and 90% are assumed. (20, 77, 103)

Statistical modelling

Statistical models are applied to explore the impact of experience variables: *operation sequence (X1), technique experience (X2),* and *age experience (X3)* on outcome. Operating surgeon is included as a random effect and treatment (six-months, twelve-months) as a fixed effect in all models.

Operation time (Y1) is analysed using a two-level multilevel linear model, which analyses the two-level data structure of patients (level 1) within surgeon (level 2).

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \cdots + u_j + \epsilon_{ij} \qquad \text{where } u_j \sim N(0, \sigma_u^2) \text{ and } \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

Equation 6

Where y_{ij} indicates the outcome for the i-th patient, β_0 is an intercept, β_1 represents treatment and ... represents additional terms added to specific models applied, specifically: s*equence (X1)* as a patient level covariate ($\beta_2 x_{ij}$) and *technique experience (X2)* and *age experience (X3)* as surgeon level covariates ($\beta_3 x_j$ and $\beta_4 x_j$ respectively).

Occurrence of fistula (Y2) is analysed using a two-level multilevel logistic model, which analyses the two-level data structure of patients (level 1) within surgeon (level 2).

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + \cdots + u_j \qquad \text{where } u_j \sim N(0, \sigma_u^2)$$

Equation 7

Where π_{ij} indicates the probability of the i-th patient having the event and other covariates are as defined in *Equation 6*.

Three models are considered for each outcome. *Model A,* which contains a treatment covariate only, represents an analysis approach ignoring any potential learning effect. *Model B* includes treatment and experience variables *operation sequence (X1)* and *technique experience (X2),* which will adjust for any potential trend due to experience gained throughout the trial and whether the surgeon had experience with the technique prior to participation. *Model C* includes treatment and experience variables *operation sequence (X1)* and *age experience (X3),* which will adjust for any potential trend due to experience gained throughout the trial and the age of the infant that the surgeon routinely operated on prior to participation. The introduction of interaction terms, between experience variables, are considered based on exploratory analysis indicating further trends.

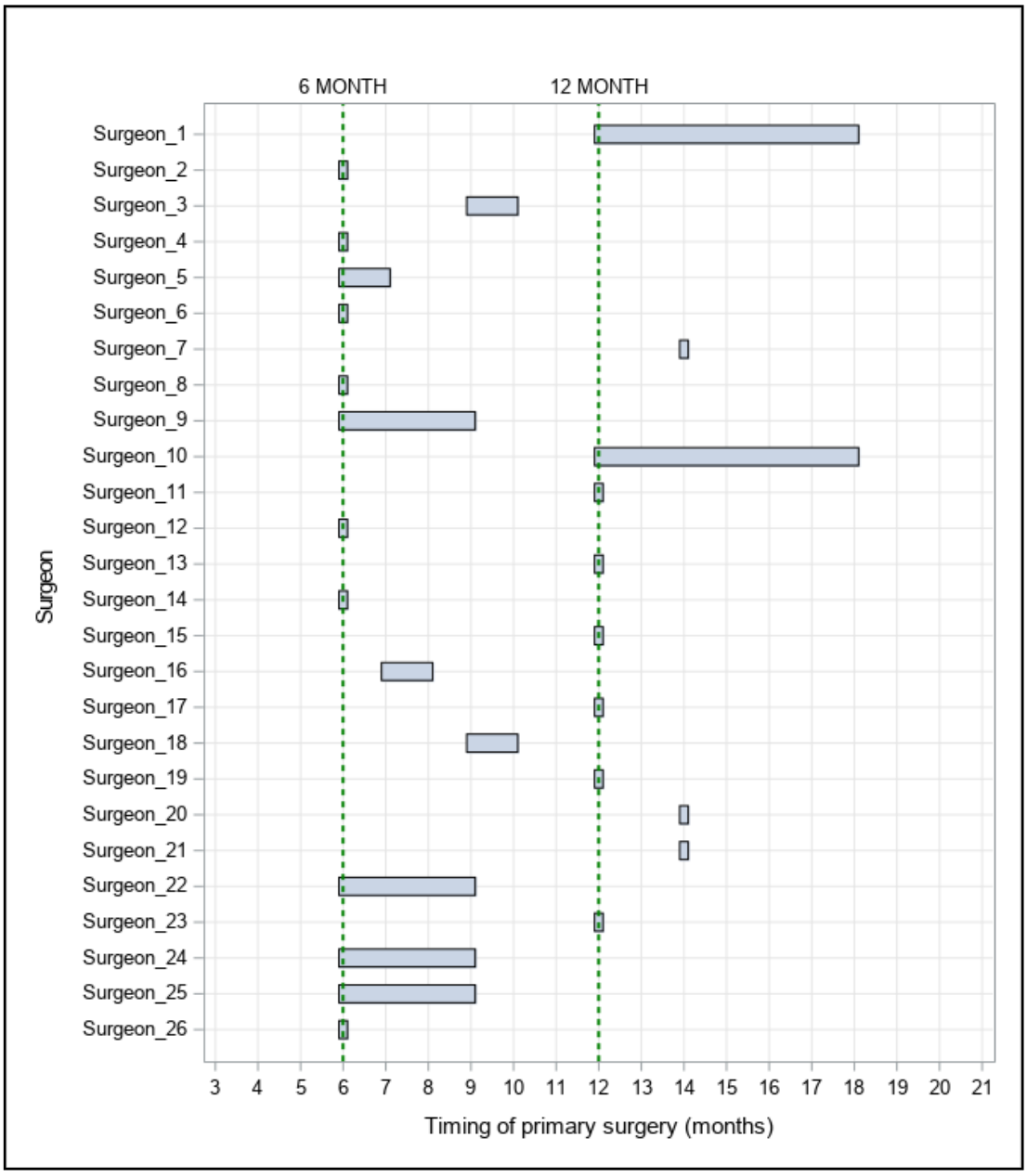
8.3. Results

8.3.1. Summary of experience

No centre had expertise in performing surgery at both six and twelve-months of age, see *Appendix Figure 1*. Twelve (54.5%) had experience of repairing clefts of infants at six-months of age and six (27.3%) had twelve-months of age. Nine (20.9%) had a preferred age range of the infant for primary cleft repair. Four (18.2%) did not include six or twelve-months of age but spanned an intermediate age, with one being later than twelve-months. Only one centre (Centre 1) had no experience in the Sommerlad technique, beyond trial specific training, prior to participation.

A total of 26 surgeons delivered the trial from the 22 centres. Surgeon 1 and Surgeon 10, as surgeons from Centre 1, had no experience in delivering the Sommerlad technique prior to participation. Two centres had two surgeons, one had three, the remaining 19 had a single operating surgeon. Six surgeons operated across two centres, although the number of operations in the second centre was comparably few compared to the overall number of trial operations. *Figure 14* and *Appendix Table 10* provides further details.

Figure 14: Timing of primary surgery for cleft palate at the time of grant application per surgeon



8.3.2. Summary of dataset

Of the 552 participants in the TOPS trial, 521 (93.5%) had surgery data and are included in this analysis. Eleven of the 26 operating surgeons, operated on at least 20 participants. The median within surgeon cluster size was 15.5. *Table 28* provides further detail.

Table 28: Summary of TOPS trial

	TOPS
Sample size	552
No. with surgery data	521 (94.4%)
No. of surgeon clusters	26
Median cluster size	15.5
No. at least 20 patients	11 (42.3%)

The trial arms were approximately equal in size, with 266 and 255 infants in the six-months and twelve-months surgery age groups respectively. Due to missing data, operation time could not be calculated for five of the 521 infants (1.0%). Occurrence of fistula follow up data were available for all infants. *Table 29* provides further detail.

Table 29: Summary of TOPS outcomes

		Six-months	Twelve-months	Overall
No. with surgery data	N	266 (51.1%)	255 (48.9%)	521
Operation time	n (%)	264 (51.2%)	252 (48.9%%)	516 (98.9%)
	Mean (SD)	86.3 (38.2)	85.9 (35.7)	84.7 (37.0)
	[Min, Max]	[30.0, 245.0]	[30.0, 210.0]	[30.0, 245.0]
	Missing	2 (0.8%)	3 (1.2%)	5 (1.0%)
Fistula				
Yes	n (%)	40 (15.0%)	33 (12.9%)	73 (14.0%)
No	n (%)	186 (85.0%)	222 (87.1%)	448 (86.0%)

8.3.3. Operation time: an outcome of surgical process

Investigating learning using visual methods

Between surgeon differences

Operation time, split by surgeon and timing of surgery, is summarised for the 516 infants with complete outcome data in *Appendix Table 11*. The distribution of the operation times by surgeon in general showed little skew, a log transformation was therefore not applied and means are reported to support consistency with existing methods later used for exploring the surgical learning curve, see *Appendix Figure 2*.

A box plot of operation time for each surgeon is provided in *Figure 15*. Within the plot, surgeons are ordered according to the number of operations each had performed, with the surgeon with the most trial operations (Surgeon 10: 85 operations) at the far left of the plot and those with the least trial operations (Surgeon 26: 3 operations) at the far right. This plot demonstrates how the average operation time varied substantially between surgeons. The plot does not suggest that between surgeon variability decreased with increasing number of operations.

A funnel plot (102) presenting the average operation time for each operating surgeon, against the number of trial operations performed for that surgeon, is provided in *Figure 16*. The target operation time is the overall trial observed mean operation time of 84.7 minutes. The plot further demonstrates the variability of surgeons operating times, with some surgeons taking longer to operate (Surgeons 1, 2 and 19) and others being generally quicker (Surgeons 4 and 5). Again, this plot explores between-surgeons times rather than within and no obvious link between number of operations and average operating time is demonstrated.

A box plot of operation time, split by treatment group, for each surgeon is provided in *Figure 17*. Within the plot, surgeons are ordered according to their timing experience prior to

participation, with the surgeons who routinely operated on younger infants at the far left and those who routinely operated on older infants at the far right. This plot shows little differences in the mean operation times, between treatment arms, in surgeons had earlier timing expertise although some who routinely operated on older infants were generally quicker in delivering the twelve-month treatment.

Figure 15: Box plot of overall operation time by surgeon

The diamond represents the mean. Within the box, the midline represents the median and the bottom and top edges the inter-quartile range. The whiskers represent the minimum and maximum values for surgeon. The x-axis is ordered by decreasing caseload (l-r).

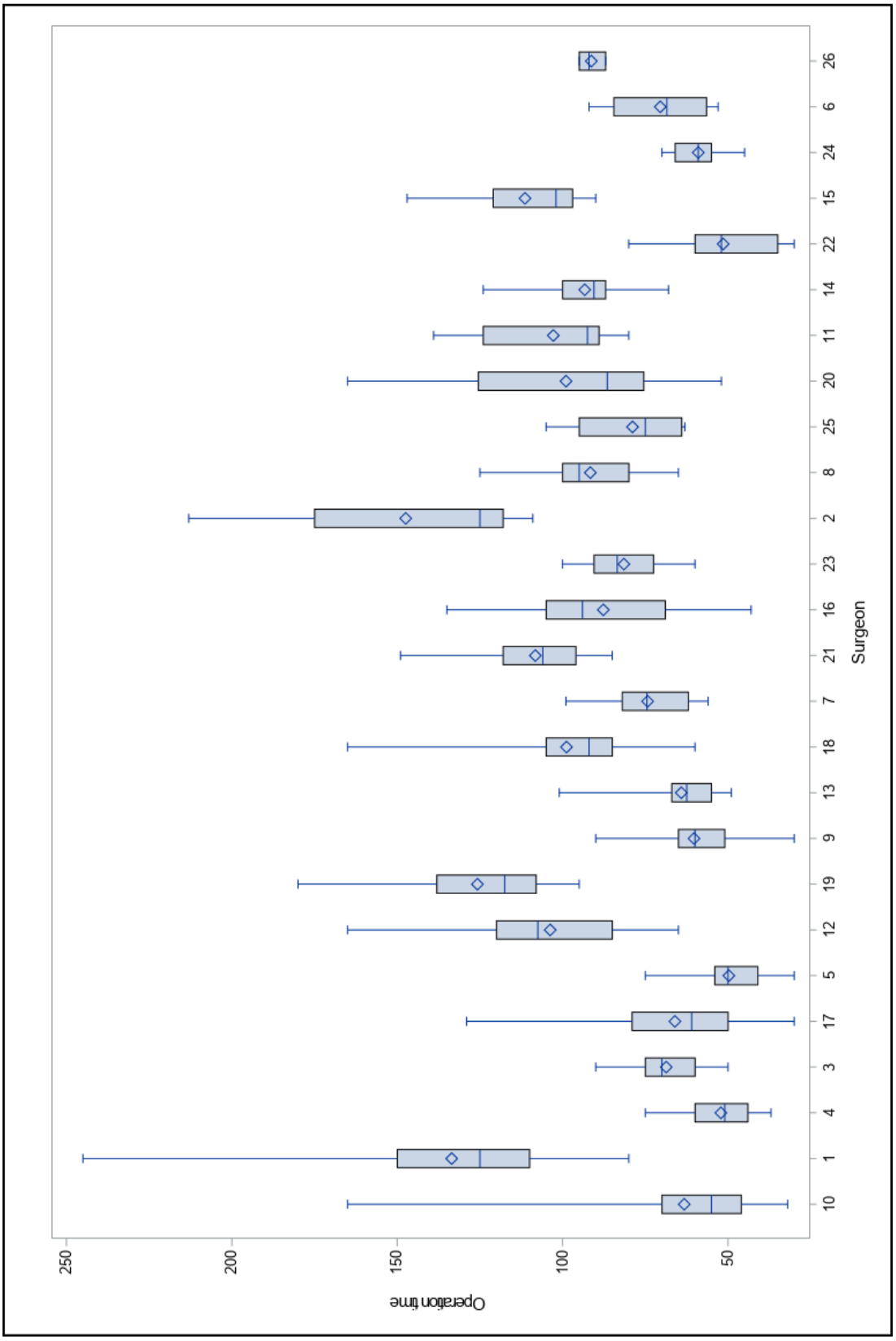


Figure 16: Funnel plot of average operation time by number of operations

The target is the overall surgeon mean operation time of 84.7 minutes. The 95% and 99.8% prediction limits around the overall mean operation time are presented. The within surgeon observed mean operation time is plotted against the number of TOPS operations for that surgeon.

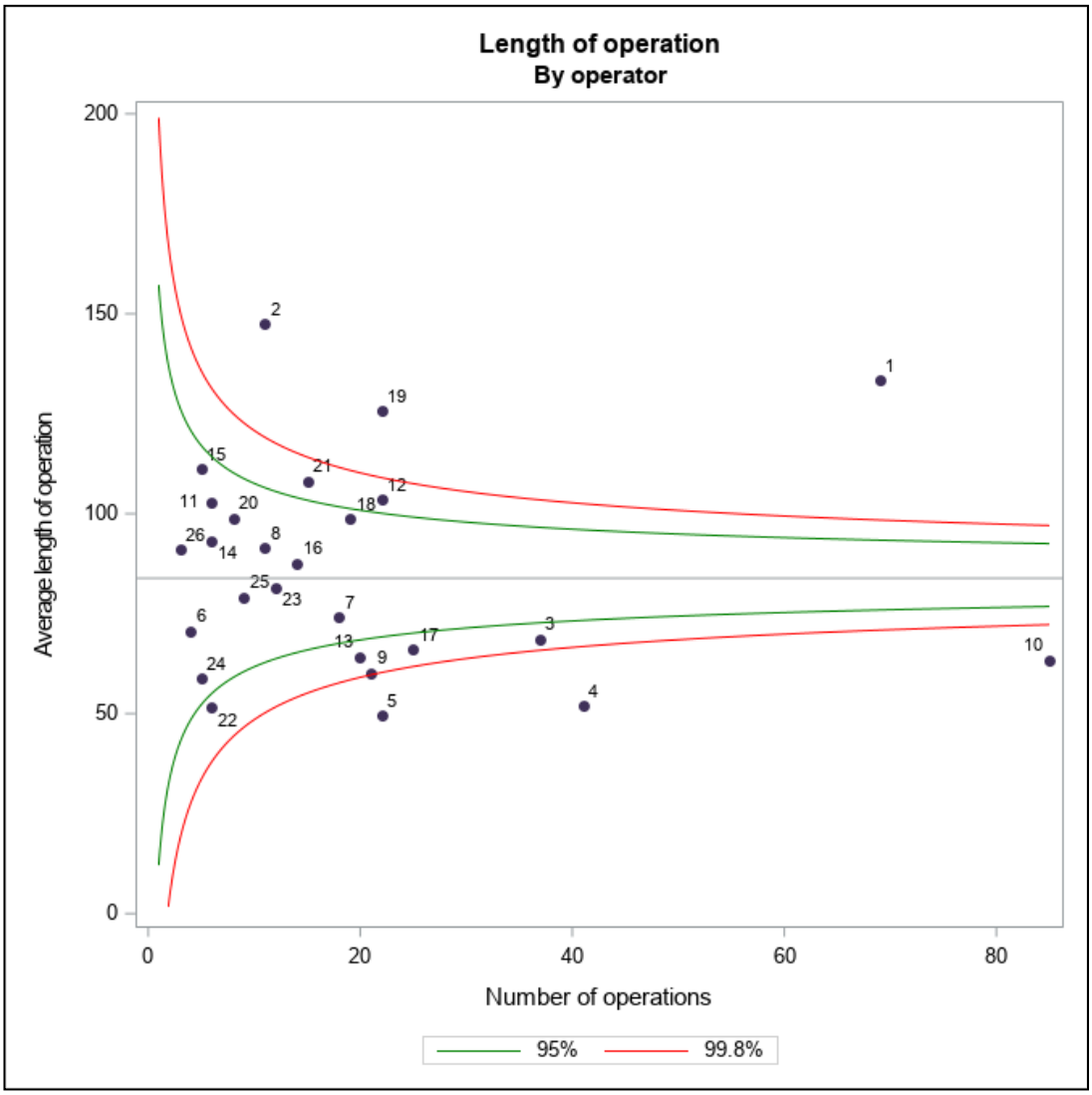
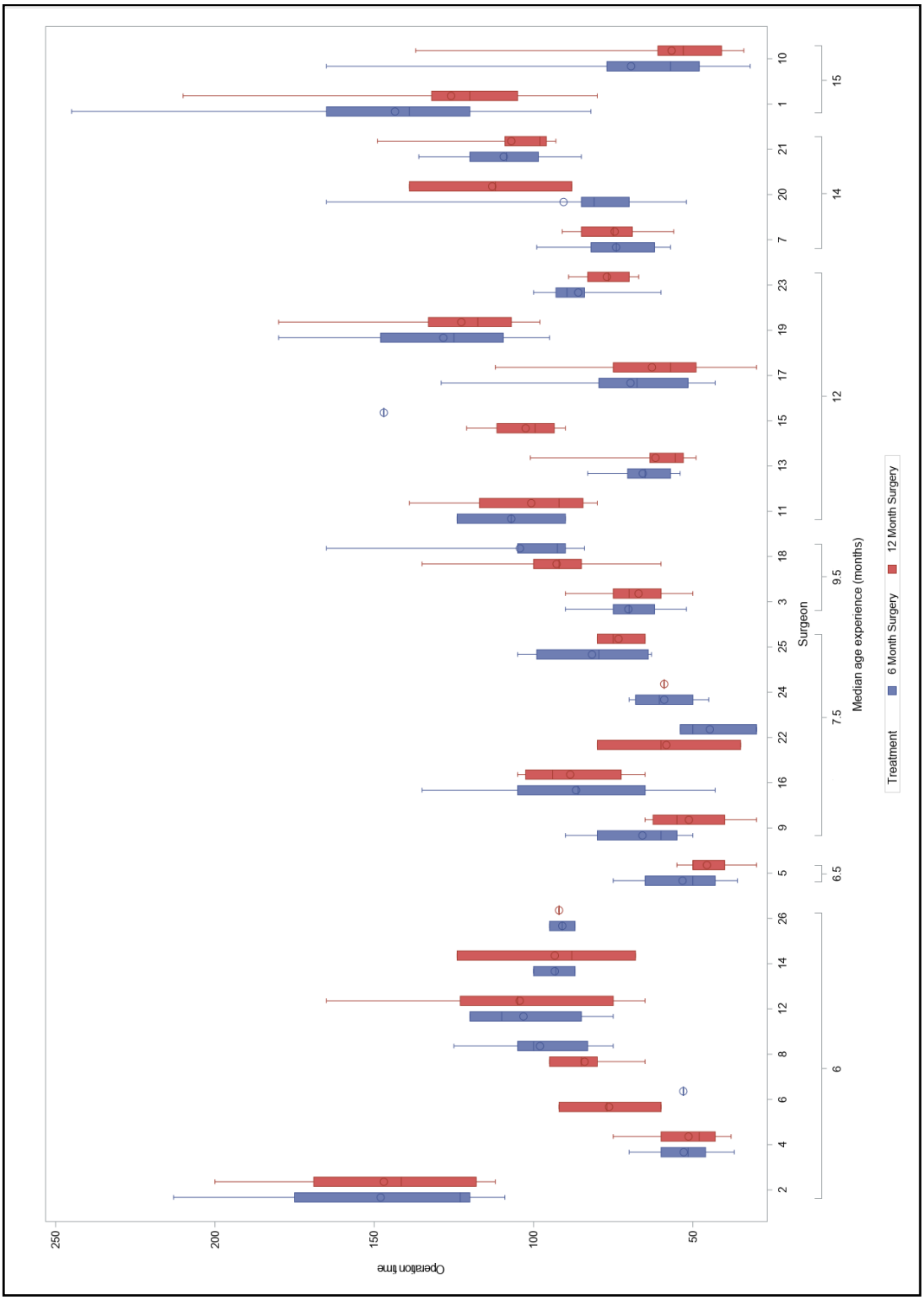


Figure 17: Box plot of operation time by treatment timing by surgeon

The circle represents the mean. Within the box, the midline represents the median and the bottom and top edges the inter-quartile range. The whiskers represent the minimum and maximum values for surgeon. The x-axis is ordered by increasing timing expertise (l-r).



Within surgeon differences

To explore within-surgeon variability, moving average plots, of order five, are presented for those surgeons who conducted at least twenty trial operations in *Figure 18* and *Appendix Figure 3.*

Surgeons 1 and 10 had no experience with the Sommerlad technique prior to participating in the trial and each delivered a greater number of operations than other participating surgeons. The moving average plots for these surgeons, see *Figure 18*, suggest a downward trend in operation time over time. Suggesting that these surgeons were getting faster at performing the technique as the trial progressed. The downward trend is evident for both treatment arms for these surgeons, despite both having experience in operating on infants aged twelve to eighteen-months primarily see *Figure 15.*

All other participating surgeons, who had experience in the Sommerlad technique, are presented in *Appendix Figure 3.* These surgeons did not show any obvious changes over time, either overall or split by treatment group. Surgeon 5, who was experienced in six to seven-months surgery prior to participation, showed a slight downward trend in operation time over time. Operation times for Surgeons 3, 4, 13 and 18, with each having varying timing experience prior to participation, were stable over time. Operation times by Surgeons 12, 17 and 19 varied yet still showed little change over the duration of the trial.

Figure 18: Moving average operation time against operation sequence

Figure presents surgeons who had no experience of the Sommerlad technique prior to participation. Moving averages of order five, based on an equal weight distribution, presented.

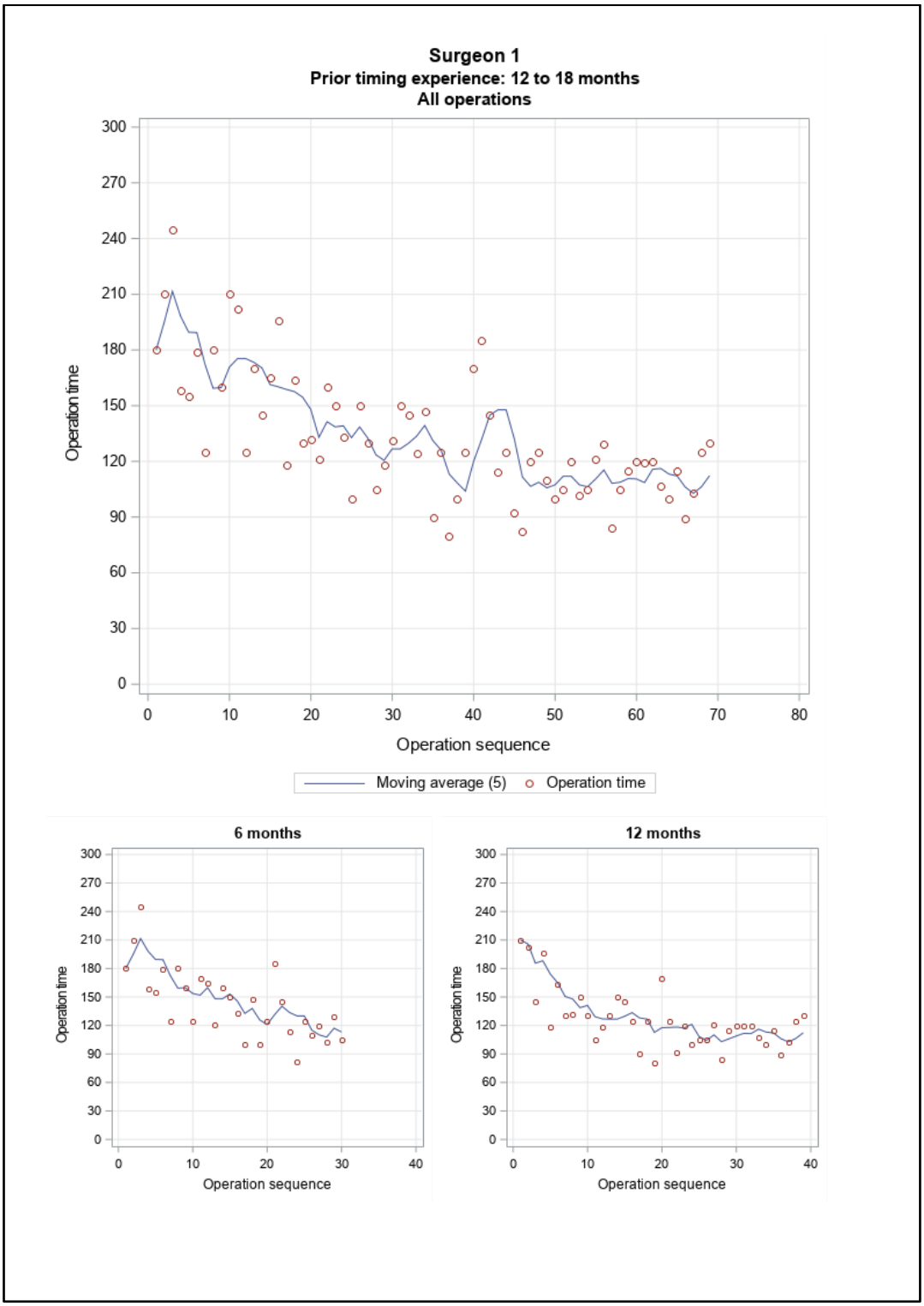
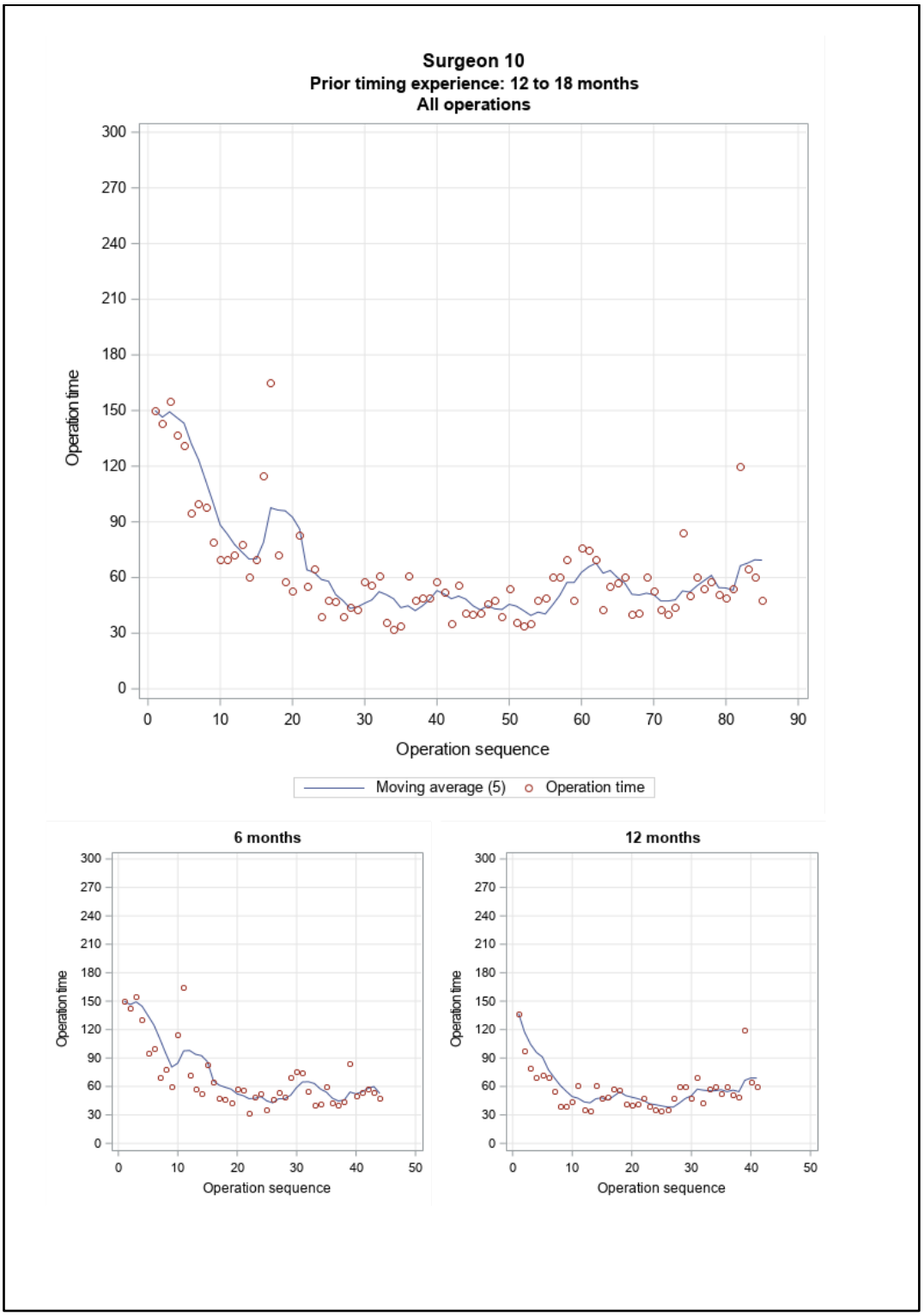


Figure 18: Moving average operation time against operation sequence

Figure presents surgeons who had no experience of the Sommerlad technique prior to participation. Moving averages of order five, based on an equal weight distribution, presented.



Investigating learning using statistical modelling

Table 30 presents the results of multilevel linear regression modelling on outcome operation time for the 516 infants with complete data.

Modal A, shows that surgery undertaken on infants at six-months of age takes approximately 7.1 minutes longer than infants at twelve-months of age ($\beta = 7.104$; 95% CI: 3.047 to 11.161; p=0.0006) after adjustment for surgeon included as a random effect.

Model B, builds on *Model A* by adjusting for sequence of operation as a learning variable and includes whether or not the surgeon had pre-trial experience of Sommerlad technique as main terms. This model indicates that the operation time decreases by approximately one minute for every two trial operations performed ($\beta = -0.568$; 95% CI: -0.709 to -0.427; p<0.0001) and that the surgery is performed 31 minutes quicker when the surgeon had prior experience with the Sommerlad technique yet this was not statistically significant ($\beta = -31.113$; 95% CI: -67.934 to 5.709; p=0.0975). When an interaction term for sequence and Sommerlad experience is incorporated into *Model B,* to further explore the trends indicated by the moving average plots (*Figure 18, Appendix Figure 3),* all variables in the model (surgical timing, experience of technique, sequence) are significant at the 0.05 level. These results further support that the slopes of the regression lines between sequence and operation time are different for surgeons with and without prior experience with the Sommerlad technique. However, the confidence intervals around the coefficient for technique experience are wide and this is reflective of the underlying distribution of operation times demonstrating wide variation between surgeons.

Model C, explores the impact of age experience on operation time by adding covariates of interest operation sequence and age experience to *Model A.* As with *Model B,* operation time decreases by approximately one minute for every two trial operations performed ($\beta =$

-0.561; 95% CI: -0.700 to -0.421; p<0.0001) and yet age experience did not alter the operating time ($\beta = 2.435$; 95% CI: -0.677 to 5.548; p=0.1249). These results indicate that whilst operation time does reduce as the number of operations increase, this difference is not influenced by age experience of each surgeon has prior to participation.

Table 30: Results of model fitting for outcome operation time (Y1)

Reference level for treatment: Twelve-months surgery; Reference level for technical experience: No technical experience.

Variables	Coefficient	95% confidence interval	*p*-value
Model A			
Treatment: Six-months surgery	7.104	3.047 to 11.161	0.0006
Model B			
Treatment: Six-months surgery	4.435	0.548 to -8.321	0.0254
Operation sequence (X1)	-0.568	-0.709 to -0.427	<0.0001
Technique experience: Yes (X2)	-31.113	-67.934 to 5.709	0.0975
Model B with interaction term			
Treatment: Six-months surgery	5.310	1.506 to 9.115	0.0063
Operation sequence (X1)	-0.727	-0.878 to -0.576	<0.0001
Technique experience: Yes (X2)	-43.442	-81.802 to -5.082	0.0265
X1 * X2	0.885	1.506 to 9.115	0.0063
Model C			
Treatment: Six-months surgery	4.476	0.590 to 8.362	0.0240
Operation sequence (X1)	-0.561	-0.700 to -0.421	<0.0001
Age experience (X3)	2.435	-0.677 to 5.548	0.1249

8.3.4. Fistula: an outcome of surgical success

Investigating learning using visual methods

Between surgeon differences

Fistula rate, split by surgeon and timing of surgery is summarised for the 521 infants with complete outcome data in *Appendix Table 12.*

A caterpillar plot of rate of fistula for each operating surgeon is provided in *Figure 19*. Within the plot, surgeons are ordered according to the number of operations each had performed, with surgeon with the most trial operations (Surgeon 10: 85 operations) at the top of the plot and those with the least trial operations (Surgeon 26: 3 operations) at the bottom. This plot demonstrates how the rate of fistula varied between surgeons. All calculable 95% confidence intervals contain the average fistula rate, and while the confidence intervals are narrower in response to larger participant numbers at the top of the graph there is no indication of increased variability

A funnel plot (102) presenting the fistula rate for each operating surgeon, against the number of trial operations performed for that surgeon, is provided in *Figure 20.* The plot further demonstrates the variability of surgeon fistula rates. Surgeon 14 is the only surgeon to have a fistula rate exceeding the upper boundary of the 95% confidence interval however the number of operations within this surgeon was low (n=6) in comparison to other trial surgeons.

As with operation time above, these exploratory plots explore between-surgeon rates rather than within surgeon rates. At this stage there is no obvious link between number of operations and a reduced fistula rate.

Figure 19: Caterpillar plot of occurrence of fistula by surgeon

Within the plot, the horizontal lines represent the ninety-five per cent confidence intervals around the diamond representing the within surgeon fistula rate. The vertical line represents the overall fistula rate observed: 14.0%.

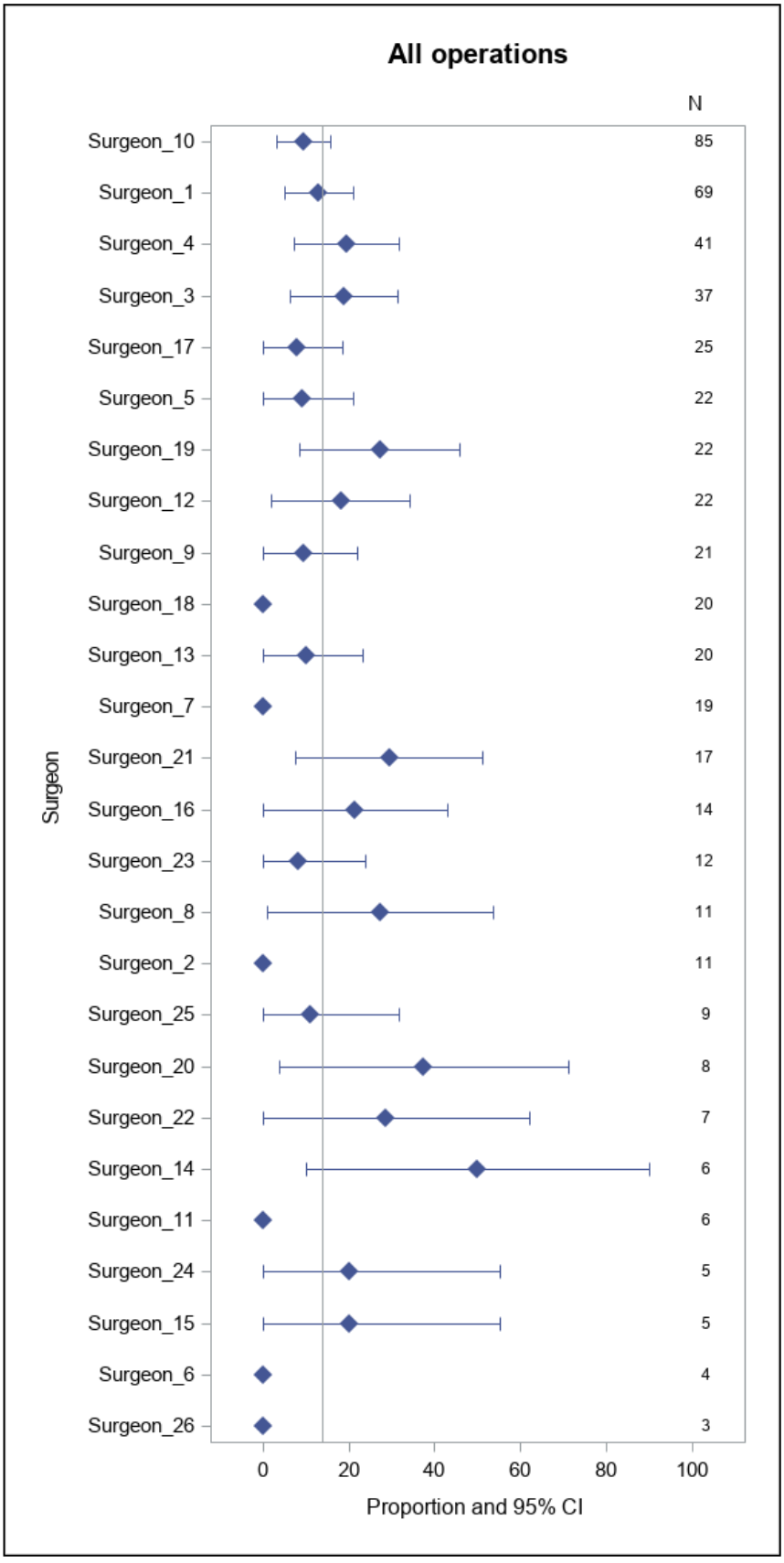
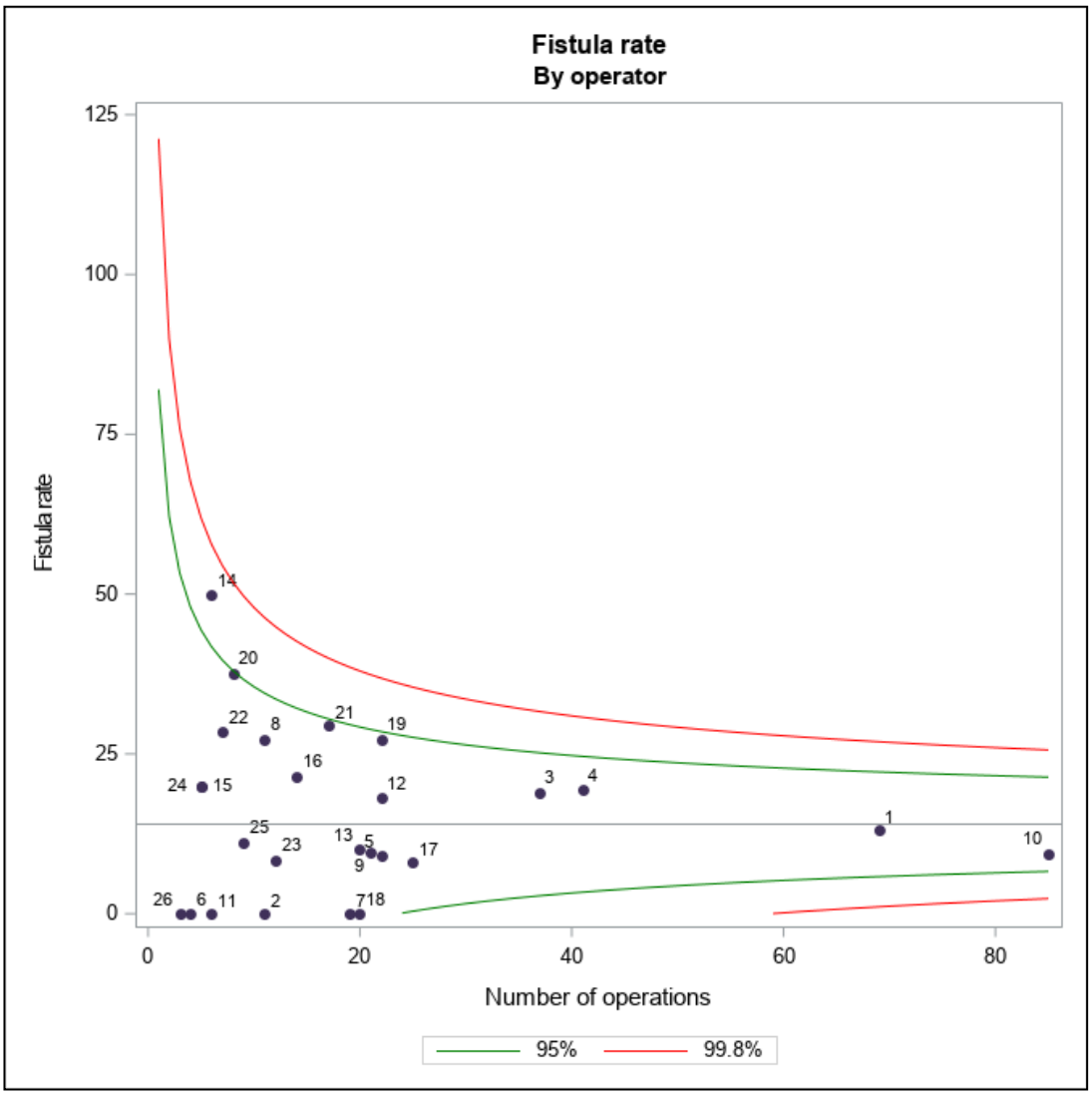


Figure 20: Funnel plot of fistula by number of operations

The target is the overall surgeon fistula rate of 14.0%. The 95% and 99.8% prediction limits around the overall fistula rate are presented. The within surgeon observed fistula rate is plotted against the number of TOPS operations for that surgeon.



Within surgeon differences

To explore within-surgeon variability, cusum plots, assuming proficiency levels 85% and 90%, are presented for surgeons who conducted at least twenty trial operations in *Figure 21* and *Appendix Figure 4* are presented to explore trend in fistula rates trends over time. When interpreting the cusum plots, a surgeon operating at the defined proficiency level (85%: blue; 90%: red) during the trial has a flat cusum line, the cusum line for a surgeon exceeding the

proficiency level decreases and increases when a surgeon operates at a rate lower than the proficiency level.

Surgeons 1 and 10 had no experience with the Sommerlad technique prior to participating in the trial and each delivered more operations than other participating surgeons. For Surgeon 1, the cusum plot indicates an initial period to approximately 15 operations where the surgeon is operating worse than the proficiency rate then the curve plateaus indicating that the proficiency rate is maintained for the remainder of the trial. However, the cusum plot for Surgeon 10 shows an initial period of approximately 35 operations where the surgeon is operating better than the proficiency rate, a period of steep deterioration in outcomes between operations 35 to 50, ending with the final 35 trial operations with a period of improvement once again. The plots split by treatment group indicate that the period of poorer outcomes tend to be within the 12 months surgery group, despite the surgeon having experience in delivering surgery to infants aged 12 to 18 months prior to participation, see *Figure 21*.

All other participating surgeons, who had prior experience in the Sommerlad technique, are presented in *Appendix Figure 4.* As with operation time, there were no clear trends within the cusum plots for these surgeons. Surgeons 5, 9, 13 and 17, each with differing background experience, showed little change in fistula rate during the trial. Surgeons 3, 4 and 12 had periods of a poor rate of fistula alongside periods of no change, with the first ten operations being a period of poor rates for Surgeon 4 and the last ten for Surgeon 3. Fistula rates for Surgeon 18 appeared to improve throughout the trial, whereas Surgeon 19 rates appeared to get worse.

Figure 21: Cusum charts for occurrence of fistula against operation sequence

Figure presented surgeons with no experience of the Sommerlad technique prior to participation. The y-axis of each figure represents the cusum score as defined in *Equation 5*.

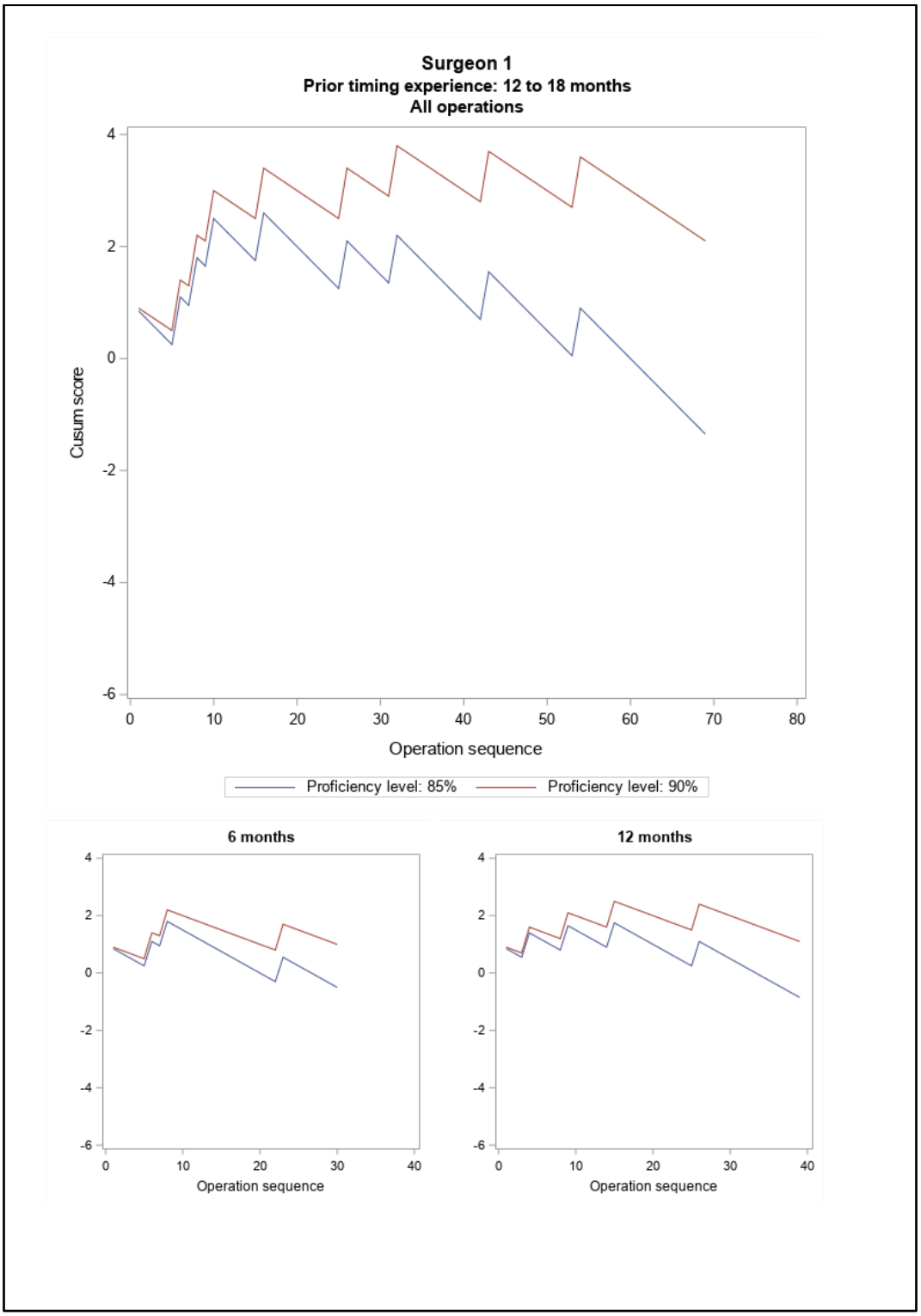
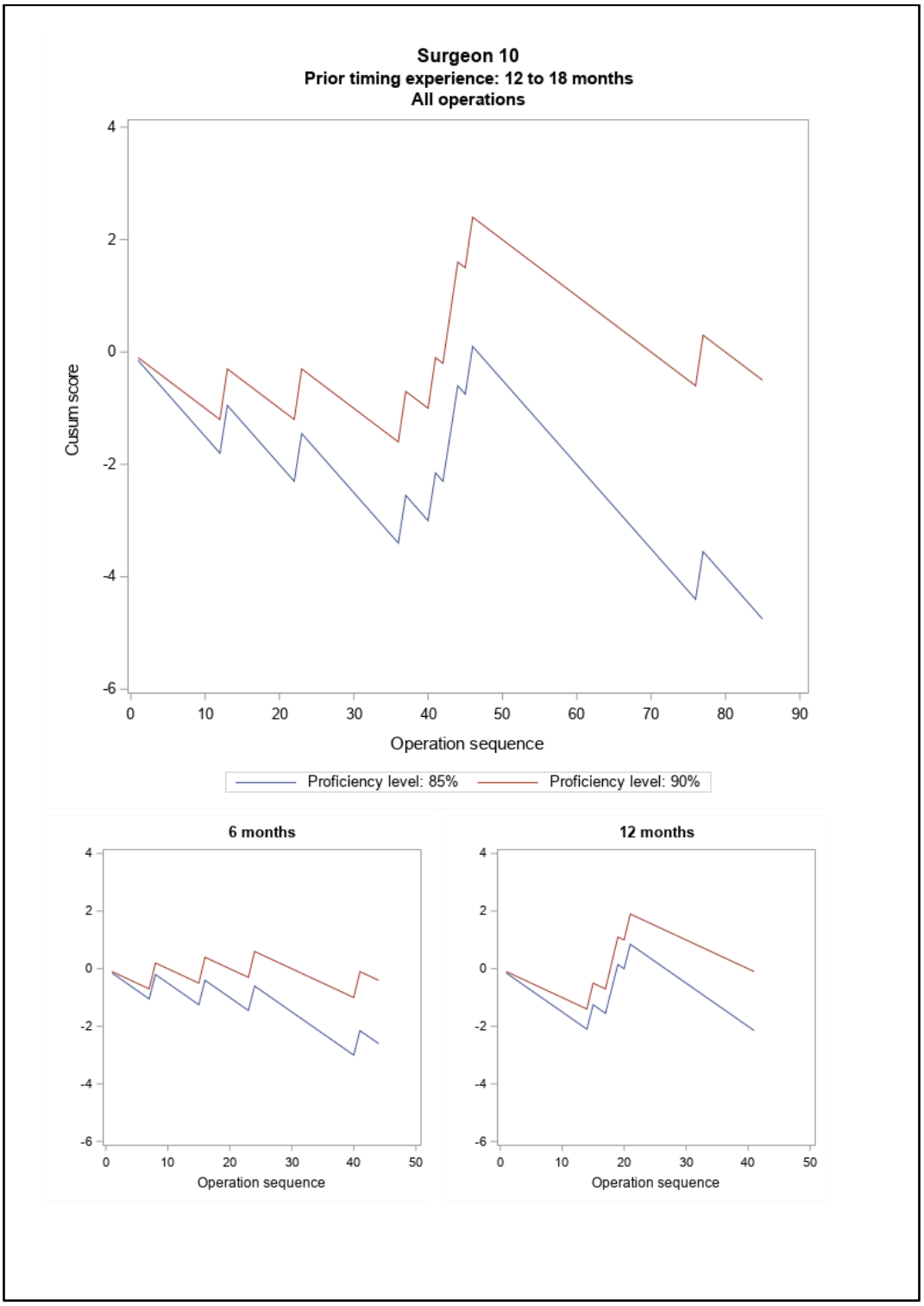


Figure 21: Cusum charts for occurrence of fistula against operation sequence

Figure presented surgeons with no experience of the Sommerlad technique prior to participation. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



Investigating learning using statistical modelling

Table 31 presents the results of multilevel logistic regression modelling on outcome occurrence of fistula for the 521 infants with complete data.

Model A, shows that there was no significant difference between the fistula rates within the two surgery groups, although the confidence intervals are wide (OR: 1.187; 95% CI: 0.720 to 1.957; p=0.5008), after adjustment for surgeon included as a random effect.

Model B, builds on *Model A* by adjusting for sequence of operation as a learning variable and includes whether or not the surgeon had pre-trial experience of Sommerlad technique as main terms. This model indicates that that neither covariate is a predictive factor of fistula (Operation sequence (X1): OR: 0.991; 95% CI: 0.441 to 2.822; p=0.3567, Technique experience (X2): OR: 1.115; 95% CI: 0.441 to 2.822; p=0.8177). These results indicate that there does not appear to be a change in fistula rate over time during the trial, and supports the cusum plots, see *Figure 21* and *Appendix Figure 4*, that indicate no trend and so further exploration of interaction terms was not warranted.

Model C, explores the impact of age experience on occurrence of fistula by adding covariates of interest operation sequence and age experience to *Model A*. Operation sequence and age experience are not statistically significant prognostic factors of fistula (Operation sequence (X1): OR: 0.991; 95% CI: 0.974 to 1.008; p=0.2914, Age experience (X3): OR: 0.984; 95% CI: 0.898 to 1.882; p=0.7249). As with the operation time analysis, prior surgical timing experience does not appear to impact fistula rate within the trial.

Table 31: Results of model fitting for outcome occurrence of fistula (Y2)

Reference level for treatment: Twelve-months surgery; Reference level for technical experience: No technical experience.

Variables	Odds ratio	95% confidence interval	*p*-value
Model A			
Treatment: Six-months surgery	1.187	0.720 to 1.957	0.5008
Model B			
Treatment: Six-months surgery	1.135	0.683 to 1.886	0.6245
Operation sequence (X1)	0.991	0.441 to 2.822	0.3567
Technique experience: Yes (X2)	1.115	0.441 to 2.822	0.8177
Model C			
Treatment: Six-months surgery	1.134	0.683 to 1.882	0.6269
Operation sequence (X1)	0.991	0.974 to 1.008	0.2914
Age experience (X3)	0.984	0.898 to 1.882	0.7249

8.4. Discussion

This chapter demonstrates approaches for analysing the presence of learning, across two different types of outcome, within a practical example of a surgical RCT: the TOPS trial.

TOPS compared surgery for primary cleft palate repair using the Sommerlad technique when infants were aged six-months or twelve-months. 521 children received surgery from 26 surgeons. Prior to participating in the trial, routine practice of infant age at surgery varied surgeon to surgeon from delivery at six-months to as old as eighteen-months. No surgeon had expertise in both surgical timings. Two surgeons had no prior experience of the Sommerlad technique. Randomisation was balanced on operating surgeon and all surgeons were calibrated on the technique of surgery prior to participation. The potential for learning could come from two sources due to varying experience base across the operating surgeons: technique experience, in using the Sommerlad technique, and age experience, routine age of

patient at surgery. Presence of learning, and any impact of this, was explored for two outcomes: operation time as a measure of surgical process and occurrence of fistula as a patient outcome.

There was evidence of a learning curve within the process outcome of surgical time. Surgeons with technique experience were faster at operating and demonstrated little change in their delivery throughout the course of the trial. Surgeons with no technique experience were initially slower but got quicker as the trial progressed, and the rate varied by treatment arm. The two surgeons, Surgeon 1 and Surgeon 10, who had no experience using the technique both had expertise in operating on infants aged 12 to 18 months prior to the study. Both surgeons achieved a plateau within both treatment timings by approximately 20 operations. Importantly, these changes did not appear to affect the formal trial outcome of fistula rate, which was also investigated within this exploration as patient outcome, where the groups were comparable before and after adjusting for sequence and Sommerlad experience.

Age experience, in terms of the infant's age that the surgeon delivered primary surgery, did not affect operation time or occurrence of fistula. Aside from Surgeon 1 and 10, the exploratory moving average and cusum plots do not show any differences between surgeons that with age experience at six-months then twelve, and this finding was further supported using statistical models.

The analysis of the TOPS data presented within this chapter has the following limitations:

1. The dataset does not contain non-randomised cases meaning that the dataset does not have information on every procedure that a surgeon has performed. In trials, unless the procedure is limited to the trial, the assessment is complicated by incomplete case series for the surgeons participating. This will be an issue across all studies, it is impractical and unethical to collect additional data on the operations of patients not consenting to trial participation.

2. The number of procedures overall, or recently according to a pre-defined time period, undertaken by operating surgeons prior to participation was not collected. This would have provided a stronger basis on which to measure experience.
3. Infant age at primary surgery was the intervention, yet the order of operation was the experience covariate in this exploration and so more six-months operations will have been performed prior to the twelve-months operations. Therefore, by trial design, any presence of a six-months learning curve may plateau before the twelve-months arm.
4. When interpreting the impact of technique experience on outcomes, the surgeons with no prior experience were from the same unit. Therefore, differences observed may be due to different protocols rather than differences due to surgical skill. The two surgeons in particular had very similar fistula rates, despite their average operating time differing by approximately one hour.
5. The number of trial operations varied from surgeon to surgeon, meaning that identifying presence of learning may not be possible in those surgeons who contributed fewer operations to the study. However, the eleven surgeons who did have a reasonable sample size, showed relatively steady outcomes over time and those contributing the most operations, Surgeon 1 and 10 who each operated on 69 and 85 operations respectively, did show evidence of a learning curve.

The limitations identified are representative of those that many statisticians will face when analysing trial data. There were no convergence issues with the modelling techniques and the exploratory nature of this investigation is emphasised. The strengths of this chapter are that it illustrates how relatively simple techniques can be applied to investigate learning curve effects. The outcomes chosen represent those identified within the literature as indicators of surgery success both within and beyond the cleft field. (17, 88, 100) However, the trends observed in operation time (as a measure of surgical process) were not mirrored in occurrence of fistula (as a measure of surgical success), raising the question of whether there is a need for better measures of learning, such as surgeon reported satisfaction with the intervention.

Surgeon experience has long been attributed as a cause of variation in cleft surgery, but until this investigation there has been no analysis of the surgical learning curve within a controlled environment to support this. (88, 89, 104) The TOPS trial, despite aiming to design out learning effects through calibration and minimising the randomisation scheme, showed some evidence that there was a change over time as surgeons became more experienced with the trial intervention. Whilst this did not appear to extend to the formal trial outcome of fistula, this investigation highlights the importance of undertaking such analyses as a precautionary measure as an aid to interpreting trial results. Particularly where trial outcomes may change over time, not only through surgeon learning, but also in trials involving other healthcare specialists where the outcome may depend on delivery, such as psychology, or where there is a systematic effect of time, such as seasonal or circadian variation. (105)

8.5. Conclusions

The TOPS trial, despite aiming to design out learning effects through calibration and randomisation, did not fully accomplish the aim of ensuring stable delivery during the study in terms of operation time, a measure of surgical process, but importantly did for fistula, a measure of surgical success and a formal outcome within the study. This investigation highlights the importance of undertaking analysis into learning as a precautionary meaurem to alleviate any concerns of surgical learning and aid trial interpretation and generalisability.

Chapter 9 : Discussion, conclusions and future work

9.1. Introduction

This chapter will provide an overview of the research that has been carried out within this thesis, as well as suggestions for future work.

The aim of this thesis has been to improve the design and analysis of randomised surgical trials. This was to be achieved by improving understanding on the impact that clustering effects, at the centre and surgeon level, and the surgical learning curve has on trial results if not adequately recognised and managed.

9.2. Research findings and conclusions

This thesis began with an overview of existing guidance for managing clustering and the surgical learning curve within the design and analysis of randomised surgical trials. This was the first review as to the extent to which existing guidance covers each of these methodological challenges. Twice the number of identified documents targeted design aspects than analysis. Most notably, no single document exists for triallists to use when designing these studies which will lead to inconsistencies in practice.

Next, a comprehensive investigation into current practice of managing clustering and learning was undertaken. As part of this investigation, a wide variety of trials, by surgical discipline, geographic location, and stage of trial development were reviewed and findings are representative of a time period of remarkable growth in the number of surgical trials being conducted.

The first review included the main trial publications of 247 randomised surgical trials and aimed to establish reporting standards. Trials were published within a two-year period within

sixteen leading journals. This novel assessment identified a lack of consideration for these effects. When considerations were made, methods varied and demonstrated poor adherence to established reporting guidelines. It was recommended that triallists consider learning and clustering on a trial-by-trial basis, and report methods used or justify where not used to inform interpretation of results. Using example trial scenarios, recommendations were provided about when and how to address learning and clustering.

The second review included fifty grant applications of randomised surgical trials and aimed to determine practice in planning for, and acknowledging, these effects when developing randomised surgical trials. Eligible applications were those funded within a four-year period by a leading UK funding body. Results indicated that, while considerations were underreported in main trial publications, funders and triallists alike appear to be aware of the need to manage learning and clustering, by centre and surgeon, during trial development. Furthermore, insight into the promising role of the funder as a potential driver of considerations is identified.

To complete understanding of practice, a survey was completed by 44 lead statisticians within the fifty UK CRC Registered Clinical Trial Units. The survey drew upon quotes from the existing guidance, relevant publications and used example trial scenarios. Promisingly, widespread awareness of challenges in designing and analysing multicentre trials was identified. Approaches used, and opinions on these, varied both across and within Units, which further indicated that the approaches used were dependent on the type of trial. Recommendations included the development of agreed principles to guide design and analysis across a range of realistic trial scenarios. This would encourage better consistency between triallists and ensure optimal methods are used as applicable.

The next chapter advanced the idea of trial-by-trial considerations, highlighting when and how clustering and learning may be managed within a pragmatic research setting. Trials were

selected such that they represented a variety of different settings, intervention types and surgical specialties. Four completed surgical trials, selected from the Liverpool Clinical Trials Centre portfolio, were summarised and described to demonstrate one Unit's experience and link back to the recommendations for practice provided in *Chapter 2*. Again, the importance of early statistical involvement in trial design was recommended. This is to reduce any impact of clustering or learning through design, and ensure that data is collected to support further investigations of either of these effects should the need arise.

The final two chapters demonstrates how further investigations into clustering and learning could be undertaken and the impact on trial results if unrecognised. Methods proposed will be particularly useful if concerns are raised, by the medical community, with regards to heterogeneity treatment effects once the main trial results are available. This reanalysis included two datasets, from the trial examples summarised in *Chapter 6*.

The first investigation explored the presence of clustering, by centre and surgeon, and its impact on trial conclusions. Two types of statistical model were applied to the datasets, one that adjusted for different levels of clustering, by centre and surgeon, and the other with no such adjustment, representing the approach whereby clustering is ignored. The reanalysis demonstrated little difference in the estimated treatment effects between the two types of model. However, this only applied to studies observing treatment differences and ICCs similar to that observed in the example trials. When data were simulated to represent trials observing greater treatment differences and levels of clustering, the adjusted model gave greater power and coverage than the unadjusted model. The need to account for clustering became vital as the odds ratio or ICC increased, as the true treatment difference was more likely to be missed by the unadjusted model.

The second investigation explored the presence of a surgical learning curve. Statistical methods were applied to a real trial dataset to demonstrate how learning can be identified and,

if necessary, controlled for within the trial analysis. Exploratory methods and statistical modelling were applied to two outcomes: operation time as a measure of surgical process and occurrence of fistula as a patient outcome. There was evidence of learning within surgical time, with surgeons less familiar with the intervention getting progressively quicker throughout the study. Importantly, this trend did not affect occurrence of fistula, which was one of the secondary outcomes of the trial. Whilst surgeon experience is often attributed as a cause of variation in surgical outcome, this investigation was the first of its type within the cleft field. The TOPS trial, despite taking precautions to minimise learning by design, provides an example where learning is still present. The importance of undertaking such analyses as a precautionary measure to aid interpretation of trial results is highlighted.

Both statistical investigations demonstrate the need for early consideration of the potential effects, and their impact on trial conclusions, when designing RCTs. This will improve data collection methods and ensure such investigations are integrated into the planned analysis of the trial. This will avoid concerns being raised about validity of the trial too late in the study and thus will support roll out of the interventions into routine practice.

9.3. Future work

It is clear from this research that triallists undertaking surgical RCTs need to be aware of the potential impact of clustering and learning. Throughout this thesis, the statistician has been advocated as a driver for making these considerations and results suggest that they are aware of these effects. However, a complete understanding of available guidance requires accessing multiple documents, requiring time and resources.

There is a need for existing documents to incorporate guidance on the management of clustering and learning. The IDEAL framework aligns perfectly with these findings as developed specifically for surgical trials. The framework is already widely used by surgical

triallists thereby providing a good platform for dissemination and, importantly as this work has identified, currently lacks in statistical guidance. (14, 15, 35, 37) Future work should address integrating the statistical themes discussed throughout this thesis into this framework as a priority. Statistical guidance should cater to different trial scenarios, another theme within this thesis, and have particular emphasis on analysis, for which current guidance is lacking. This would help drive standardisation, on a trial specific basis, within practice and ultimately support the delivery of surgical trials to come.

In addition, there are a number of other areas that future work should investigate. The statistical investigations in *Chapter 7* and *Chapter 8* were conducted on trials with a reasonably large sample size. In reality, trials can (and often are) much smaller. Meaningful analyses can be difficult when there are very few patients per cluster, such as in the NERVES trial in *Chapter 6*. Understanding the limitations of the methods demonstrated, and alternatives or adaption to be applicable, within smaller studies would support wider use of the methods presented. Likewise, the simulations focus on impact of surgeon, whereas further work may investigate the impact of centre, again this would open the methods used to a wider variety of trials. The investigation into learning identified that, whilst operation time did vary within surgeon during the course of the trial, this did not extend to affect the formal trial outcome of occurrence of fistula. This raises the question of whether operation time, and safety measures, are adequate measures of learning despite being commonly used within the wider surgical literature. Better measures could be developed, such as surgeon reported outcomes, and collected as part of the trial for explorations such as these.

Importantly, clustering and learning are not unique to surgery and these findings apply to the wider medical field also. The role of the surgeon in delivering a surgical trial is comparable to the role of centre in delivering treatments in drug trials. Clustering can be present when the intervention is part of a wider care bundles, which can differ due to local protocols, or where hospitals vary in specialty or demographic. In terms of learning, changes in outcome over time

extend to any other healthcare specialty where outcome may depend on delivery, such as psychology, or where there is a systematic effect of time, such as seasonal or circadian variation. Beyond surgery, data collection should be standardised so that, should the need arise, the issues of clustering and learning can be explored if the need arises. It is important therefore that future work aims to increase awareness and training of these effects beyond the surgical field, and this can be achieved through further collaboration with the UKCRC Registered Clinical Trials Units and integration of these findings into broader guidance, such as the MRC Guidelines for Complex Interventions. (13, 61, 63)

Bibliography

1. Meara JG, Leather AJ, Hagander L, Alkire BC, Alonso N, Ameh EA, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. Lancet. 2015;386(9993):569-624.
2. Shrime MG, Bickler SW, Alkire BC, Mock C. Global burden of surgical disease: an estimation from the provider perspective. Lancet Glob Health. 2015;3 Suppl 2:S8-9.
3. Horton R. Surgical research or comic opera: questions, but few answers. The Lancet. 1996;347(9007):984-5.
4. Royal College of Surgeons of England. Surgical Trials Initiative Royal College of Surgeons of England web site: Royal College of Surgeons of England; 2021 [Available from: https://www.rcseng.ac.uk/standards-and-research/research/surgical-trials-initiative/.
5. McCall B. UK implements national programme for surgical trials. Lancet. 2013;382(9898):1083-4.
6. National Institute for Health Research. Surgery: National Institute for Health Research; 2021 [Available from: https://www.nihr.ac.uk/explore-nihr/specialties/surgery.htm.
7. GlobalSurg, NIHR Global Health Research Unit on Global Surgery. Improving surgical outcomes through collaborative research. Global Surgery web site: National Institute for Health Research; 2020 [Available from: https://globalsurg.org/.
8. The Lancet. The Lancet Commission on Global Surgery The Lancet web site: The Lancet; 2015 [Available from: https://www.thelancet.com/commissions/global-surgery.
9. National Institute for Health Research. Funding and Awards 2021 [Available from: https://fundingawards.nihr.ac.uk/search.
10. National Institutes of Health. ClinicalTrials.gov ClinicalTrials.gov web site: U.S. National Library of Medicine,; 2021 [Available from: https://clinicaltrials.gov/.
11. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. J Clin Nurs. 2003;12(1):77-84.
12. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. Trials. 2009;10(1):9.
13. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions2019; 2019(January 23rd). Available from: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/.
14. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, et al. Challenges in evaluating surgical innovation. Lancet. 2009;374.
15. McCulloch P, Altman DG, Campbell B, Flum DR, Glasziou P, Marshall J, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374.
16. ICH Expert Working Group. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. Stat Med. 1999;18(15):1905-42.
17. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effec ts in surgical trials. Clin Trials. 2004;1.
18. Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. Clinical Trials. 2005;2(2):152-62.
19. Oltean H, Gagnier JJ. Use of clustering analysis in randomized controlled trials in orthopaedic surgery. BMC Med Res Methodol. 2015;15:17.
20. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. Health Technol Assess. 2001;5(12):1-79.
21. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, Grp C. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. Ann Intern Med. 2008;148(4):295-309.

22. National Institute for Health Research. Planning a Randomised Controlled Trial (RTC) – Points to Consider2016. Available from: https://www.ct-toolkit.ac.uk/routemap/trial-planning-and-design/downloads/planning-a-randomised-controlled-trial.pdf/.
23. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. Brit Med J. 2008;337.
24. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661.
25. Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, Gamble C. Randomized trials involving surgery did not routinely report considerations of learning and clustering effects. Journal of Clinical Epidemiology. 2019;107:27-35.
26. Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, Gamble C. Funders improved the management of learning and clustering effects through design and analysis of randomized trials involving surgery. Journal of Clinical Epidemiology. 2019;113:28-35.
27. Jackson DL. Reporting results of latent growth modeling and multilevel modeling analyses: some recommendations for rehabilitation psychology. Rehabil Psychol. 2010;55(3):272-85.
28. ICH Expert Working Group. ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6(R1). ICH Harmonisation for better health web site: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.; 1996 [Available from: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf.
29. Medicines & Healthcare products Regulatory Agency. Clinical investigations of medical devices – statistical considerations2021. Available from: Statistical_considerations_clinical_investigations_-_May_2021.pdf (publishing.service.gov.uk).
30. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, Group CN. 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):40-7.
31. Elias KM, Stone AB, McGinigle K, Tankou JI, Scott MJ, Fawcett WJ, et al. The Reporting on ERAS Compliance, Outcomes, and Elements Research (RECOvER) Checklist: A Joint Statement by the ERAS((R)) and ERAS((R)) USA Societies. World J Surg. 2019;43(1):1-8.
32. Vanhie A, Meuleman C, Tomassetti C, Timmerman D, D'Hoore A, Wolthuis A, et al. Consensus on Recording Deep Endometriosis Surgery: the CORDES statement. Hum Reprod. 2016;31(6):1219-23.
33. ICH Expert Working Group. ICH Harmonised Tripartite Guidelines: General Considerations for Clinical Trials E8.1998. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-8-general-considerations-clinical-trials-step-5_en.pdf.
34. ICH Expert Working Group. ICH Harmonised Tripartite Guideline: Structure and Content of Clinical Study Reports E3. ICH Harmonsation for better health web site: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.; 1995 [Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf.
35. Bilbro NA, Hirst A, Paez A, Vasey B, Pufulete M, Sedrakyan A, et al. The IDEAL Reporting Guidelines: A Delphi Consensus Statement Stage Specific Recommendations for Reporting the Evaluation of Surgical Innovation. Ann Surg. 2021;273(1):82-5.
36. Committee for Medicinal Products for Human Use (CHMP). Guideline on adjustment for baseline covariates in clinical trials2015 30th November 2021. Available from: https://www.ema.europa.eu/en/adjustment-baseline-covariates-clinical-trials.

37. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Balliol C, et al. Evaluation and stages of surgical innovations. Lancet. 2009;374(9695):1089-96.
38. Harrysson IJ, Cook J, Sirimanna P, Feldman LS, Darzi A, Aggarwal R. Systematic review of learning curves for minimally invasive abdominal surgery: a review of the methodology of data collection, depiction of outcomes, and statistical analysis. Annals of surgery. 2014;260(1):37-45.
39. Mahawar KK, Kumar G, Malviya A. Who publishes in leading general surgical journals? The divide between the developed and developing worlds. Asian Journal of Surgery. 2006;29(3):140-4.
40. Wu R, Glen P, Ramsay T, Martel G. Reporting quality of statistical methods in surgical observational studies: protocol for systematic review. Syst Rev. 2014;3:70.
41. National Institute for Health Research. Feasibility and Pilot studies National Institute for Health Research web site: National Institute for Health Research; [Available from: https://www.nihr.ac.uk/funding-and-support/documents/funding-for-research-studies/research-programmes/PGfAR/CCF-PGfAR-Feasibility-and-Pilot-studies.pdf.
42. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. Bmj-Brit Med J. 2005;330(7482):88-91.
43. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. BMJ. 2012;345.
44. Dogan S, Bayraktar C. Endoscopic tympanoplasty: learning curve for a surgeon already trained in microscopic tympanoplasty. Eur Arch Oto-Rhino-L. 2017;274(4):1853-8.
45. Wu WB, Xu J, Wen W, Yu Y, Xu XF, Zhu Q, et al. Learning curve of totally thoracoscopic pulmonary segmentectomy. Front Med-Prc. 2018;12(5):586-92.
46. Papachristofi O, Jenkins D, Sharples LD. Assessment of learning curves in complex surgical interventions: a consecutive case-series study. Trials. 2016;17(1):1-10.
47. Papachristofi O, Klein AA, Mackay J, Nashef S, Fletcher N, Sharples LD, et al. Effect of individual patient risk, centre, surgeon and anaesthetist on length of stay in hospital after cardiac surgery: Association of Cardiothoracic Anaesthesia and Critical Care (ACTACC) consecutive cases series study of 10 UK specialist centres. Bmj Open. 2017;7(9).
48. Shantikumar S, Wigley J, Hameed W, Handa A. A survey of instructions to authors in surgical journals on reporting by CONSORT and PRISMA. Ann Roy Coll Surg. 2012;94(7):468-71.
49. National Institute for Health Research. Health Technology Assessment National Institute for Health Research web site: National Institute for Health Research; [Available from: https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technology-assessment/.
50. National Institute for Health Research. Efficacy and Mechanism Evaluation National Institute for Health Research web site: National Institute for Health Research; [Available from: https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/efficacy-and-mechanism-evaluation/.
51. National Institute for Health Research. National Institute for Health Research Journals Library National Institute for Health Research web site: National Institute for Health Research; [Available from: https://www.journalslibrary.nihr.ac.uk/journals/.
52. Gsponer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, Schmidli H. A practical guide to Bayesian group sequential designs. Pharm Stat. 2014;13(1):71-80.
53. National Institute for Health Research. Clinical Trials Toolkit: Funding proposal National Institute for Health Research web site: National Institute for Health Research; [Available from: http://www.ct-toolkit.ac.uk/routemap/funding-proposal/.
54. Blencowe NS, Mills N, Cook JA, Donovan JL, Rogers CA, Whiting P, et al. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. Br J Surg. 2016;103(10):1377-84.
55. McCulloch P, Feinberg J, Philippou Y, Kolias A, Kehoe S, Lancaster G, et al. Progress in clinical research in surgery and IDEAL. Lancet. 2018;392(10141):88-94.

56. Conroy EJ, Blazeby JM, Burnside G, Cook JA, Gamble C. Managing clustering effects and learning effects in the design and analysis of multicentre randomised trials: a survey to establish current practice. Trials. 2020;21(1):433.
57. UKCRC Registration ID Numbers United Kingdom Clinical Research Collaboration web site: United Kingdom Clinical Research Collaboration; 2018 [Available from: https://cdn.ymaws.com/www.ukcrc-ctu.org.uk/resource/resmgr/registration_ids/2018-19_reg_ids_oct18.pdf. .
58. Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. Trials. 2015;16:405.
59. Biggs K, Hind D, Gossage-Worrall R, Sprange K, White D, Wright J, et al. Challenges in the design, planning and implementation of trials evaluating group interventions. Trials. 2020;21(1):116.
60. Cook JA, Campbell M, Gillis K, Skea Z. Surgeons' and methodologists' perceptions of utilising an expertise based randomised controlled trial design: a qualitative study. . Trials. 2018;19(478).
61. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337.
62. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350:h2147.
63. McFadden E, Bashir S, Canham S, Darbyshire J, Davidson P, Day S, et al. The impact of registration of clinical trials units: the UK experience. Clin Trials. 2015;12(2):166-73.
64. Cousins S, Blencowe NS, Blazeby JM. What is an invasive procedure? A definition to inform study design, evidence synthesis and research tracking. BMJ Open. 2019;9(7):e028576.
65. Liverpool Clinical Trials Centre. Research 2020 [Available from: https://lctc.org.uk/research?portfolio=Surgical.
66. Richards HK, Seeley HM, Pickard JD. Efficacy of antibiotic-impregnated shunt catheters in reducing shunt infection: data from the United Kingdom Shunt Registry. J Neurosurg Pediatr. 2009;4(4):389-93.
67. Jenkinson MD, Gamble C, Hartley JC, Hickey H, Hughes D, Blundell M, et al. The British antibiotic and silver-impregnated catheters for ventriculoperitoneal shunts multi-centre randomised controlled trial (the BASICS trial): study protocol. Trials. 2014;15:4.
68. Mallucci CL, Jenkinson MD, Conroy EJ, Hartley JC, Brown M, Dalton J, et al. Antibiotic or silver versus standard ventriculoperitoneal shunts (BASICS): a multicentre, single-blinded, randomised trial and economic evaluation. Lancet. 2019;394(10208):1530-9.
69. Mallucci CL, Jenkinson MD, Conroy EJ, Hartley JC, Brown M, Moitt T, et al. Silver-impregnated, antibiotic-impregnated or non-impregnated ventriculoperitoneal shunts to prevent shunt infection: the BASICS three-arm RCT. Health Technol Assess. 2020;24(17):1-114.
70. Wilby MJ, Hopkins C, Bedson E, Howlin S, Burnside G, Conroy EJ, et al. Nerve root block versus surgery (NERVES) for the treatment of radicular pain secondary to a prolapsed intervertebral disc herniation: study protocol for a multi-centre randomised controlled trial. Trials. 2018;19(1):475.
71. Wilby MJ, Best A, Wood E, Burnside G, Bedson E, Short H, et al. Surgical microdiscectomy versus transforaminal epidural steroid injection in patients with sciatica secondary to herniated lumbar disc (NERVES): a phase 3, multicentre, open-label, randomised controlled trial and economic evaluation. Lancet Rheumatol. 2021;3(5):e347-e56.
72. Wilby MJ, Best A, Wood E, Burnside G, Bedson E, Short H, et al. Microdiscectomy compared with transforaminal epidural steroid injection for persistent radicular pain caused by prolapsed intervertebral disc: the NERVES RCT. Health Technol Assess. 2021;25(24):1-86.
73. Kahan BC, Morris TP. Assessing potential sources of clustering in individually randomised trials. BMC Med Res Methodol. 2013;13:58.

74. Shaw W, Semb G, Lohmander A, Persson C, Willadsen E, Clayton-Smith J, et al. Timing Of Primary Surgery for cleft palate (TOPS): protocol for a randomised trial of palate surgery at 6 months versus 12 months of age. BMJ Open. 2019;9(7):e029780.
75. Peterson-Falzone S. Optimal Age for Palatoplasty to Facilitate Normal Speech Development: What is the Evidence?: Springer, Berlin, Heidelberg; 2006.
76. Conroy EJ, Cooper R, Shaw W, Persson C, Willadsen E, Munro KJ, et al. A randomised controlled trial comparing palate surgery at 6 months versus 12 months of age (the TOPS trial): a statistical analysis plan. Trials. 2021;22(1):5.
77. Sommerlad BC. A technique for cleft palate repair. Plast Reconstr Surg. 2003;112(6):1542-8.
78. Halloran CM, Platt K, Gerard A, Polydoros F, O'Reilly DA, Gomez D, et al. PANasta Trial; Cattell Warren versus Blumgart techniques of panreatico-jejunostomy following pancreato-duodenectomy: Study protocol for a randomized controlled trial. Trials. 2016;17:30.
79. National Health Service. Treatment: Pancreatic cancer. 2018 [Available from: https://www.nhs.uk/conditions/pancreatic-cancer/treatment/.
80. Halloran C, Neoptolemos JP, Platt K, Jackson R, Reddy S, Oreilly D, et al. PANasta Trial: Cattell Warren versus Blumgart techniques of pancreatico-jejunostomy following pancreato-duodenectomy—A double-blinded multi-centered trial, trial results. Journal of Clinical Oncology. 2020;38(15_suppl):4619-.
81. Liverpool Clinical Trials Centre. PANASTA 2020 [Available from: https://lctc.org.uk/research/panasta/.
82. Alsagheir A, Koziarz A, Belley-Cote EP, Whitlock RP. Expertise-based design in surgical trials: a narrative review. Can J Surg. 2021;64(6):E594-E602.
83. Walter SD, Ismaila AS, Devereaux PJ, Investigators SS. Statistical issues in the design and analysis of expertise-based randomized clinical trials. Stat Med. 2008;27(30):6583-96.
84. Walters SJ. Therapist effects in randomised controlled trials: what to do about them. J Clin Nurs. 2010;19(7-8):1102-12.
85. Cook JA, Bruckner T, MacLennan GS, Seiler CM. Clustering in surgical trials-database of intracluster correlations. Trials. 2012;13(1):2.
86. Campbell MJ, Machin D, Walters SJ. Medical statistics: a textbook for the health sciences: John Wiley & Sons; 2010.
87. Lee KJ, Thompson SG. Clustering by health professional in individually randomised trials. BMJ. 2005;330(7483):142-4.
88. Rautio J, Andersen M, Bolund S, Hukki J, Vindenes H, Davenport P, et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 2. Surgical results. J Plast Surg Hand Surg. 2017;51(1):14-20.
89. Shaw W, Semb G. The Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 11. What next? J Plast Surg Hand Surg. 2017;51(1):88-93.
90. Flight L, Allison A, Dimairo M, Lee E, Mandefield L, Walters SJ. Recommendations for the analysis of individually randomised controlled trials with clustering in one arm - a case of continuous outcomes. BMC Med Res Methodol. 2016;16(1):165.
91. Leckie G, French R. Three-Level Multilevel Models – MLwiN Practical. Multilevel modelling online course [Internet]. 2013:[1-56 pp.]. Available from: http://www.bristol.ac.uk/cmm/learning/course.html.
92. Kiernan K. Insights into Using the GLIMMIX Procedure to Model Categorical Outcomes with Random Effects2018. Available from: https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2179-2018.pdf.
93. Austin PC, Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. Stat Med. 2017;36(20):3257-77.
94. Goldstein H. Multilevel statistical models: John Wiley & Sons; 2011.
95. Chen H, Cohen P, Chen S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. Communications in Statistics—simulation and Computation®. 2010;39(4):860-4.

96. Gibbs P, Kiernan K. Simulating Data for Complex Linear Models. SAS Global Forum 2020 [Internet]. 2020. Available from: https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2020/4400-2020.pdf.
97. Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how? BMC Med Res Methodol. 2014;14:20.
98. Chu R, Thabane L, Ma J, Holbrook A, Pullenayegum E, Devereaux PJ. Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: a simulation study. BMC Med Res Methodol. 2011;11:21.
99. Corrigan N, Marshall H, Croft J, Copeland J, Jayne D, Brown J. Exploring and adjusting for potential learning effects in ROLARR: a randomised controlled trial comparing robotic-assisted vs. standard laparoscopic surgery for rectal cancer resection. Trials. 2018;19(1):339.
100. Losee JE, Kirschner RE, Smith DM, Straub AE, Lawrence C. Comprehensive Cleft Care: Family Edition: CRC Press/Taylor & Francis Group; 2015.
101. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. Western Journal of Emergency Medicine. 2017;18(6):1075.
102. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med. 2005;24(8):1185-202.
103. Smyth AG, Wu J. Cleft Palate Outcomes and Prognostic Impact of Palatal Fistula on Subsequent Velopharyngeal Function-A Retrospective Cohort Study. Cleft Palate Craniofac J. 2019;56(8):1008-12.
104. Butterworth S, Hodgkinson EL, Sainsbury DCG, Hodgkinson PD. Surgical Variation Between Consultant Surgeons Performing a Sommerlad Radical Intravelar Veloplasty. Cleft Palate Craniofac J. 2021;58(12):1490-9.
105. Altman DG, Royston JP. The hidden effect of time. Stat Med. 1988;7(6):629-37.

Appendix 1 – Published papers

Publications from this thesis

Work from Chapters 3, 4 and 5 has been published as research articles and conference papers. Full references for the relevant articles are shown and copies included.

Chapter 3

Research article

Randomised trials involving surgery do not report considerations of learning and clustering effects.

EJ Conroy[1], A Rosala-Hallas[1], JM Blazeby[2], G Burnside[1], JA Cook[3], C Gamble[1,4]

[1] Department of Biostatistics, University of Liverpool, Liverpool, UK; [2] Centre for Surgical Research, Population Health Sciences, University of 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.

Journal of Clinical Epidemiology. 2019; 107: 27 – 35.

Conference paper

Meeting abstracts from the 4th International Clinical Trials Methodology Conference (ICTMC 2017):

Design and analysis of the learning curve and clustering effects in randomised surgical trials.

EJ Conroy[1], JM Blazeby[2], G Burnside[1], JA Cook[3], C Gamble[1,4]

[1] Department of Biostatistics, University of Liverpool, Liverpool, UK; [2] Centre for Surgical Research, Population Health Sciences, University of 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.

Trials. 2017. 18 (Suppl 1):200 - P129.

Chapter 4

Research article

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

EJ Conroy[1], A Rosala-Hallas[1], JM Blazeby[2], G Burnside[1], JA Cook[3], C Gamble[1,4]

[1] Department of Biostatistics, University of Liverpool, Liverpool, UK; [2] Centre for Surgical Research, Population Health Sciences, University of 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.

Journal of Clinical Epidemiology. 2019; 113: 28-35.

Conference paper

Meeting abstracts from the 5th International Clinical Trials Methodology Conference (ICTMC 2019):

Managing learning and clustering effects at the design stage in randomised surgical trials: a retrospective cohort of trial funding applications.

EJ Conroy[1], A Rosala-Hallas[1], JM Blazeby[2], G Burnside[1], JA Cook[3], C Gamble[1,4]

[1] Department of Biostatistics, University of Liverpool, Liverpool, UK; [2] Centre for Surgical Research, Population Health Sciences, University of 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.

Trials. 2019. 20 (Suppl 1): 579 -P-20.

Chapter 5

Research article

Managing clustering effects and learning effects in the design and analysis of multicentre randomised trials: A survey to establish current practice.

EJ Conroy[1], JM Blazeby[2], G Burnside[1], JA Cook[3], C Gamble[1,4]

[1] Department of Biostatistics, University of Liverpool, Liverpool, UK; [2] Centre for Surgical Research, Population Health Sciences, University of 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.

Trials. 2020. 21, 433.

Conference paper

Meeting abstracts from the 5th International Clinical Trials Methodology Conference (ICTMC 2019):

Allowance for learning and clustering effects in the design and analysis of multicentre randomised trials: current practice and experiences.

EJ Conroy[1], JM Blazeby[2], G Burnside[1], JA Cook[3], C Gamble[1,4]

[1] Department of Biostatistics, University of Liverpool, Liverpool, UK; [2] Centre for Surgical Research, Population Health Sciences, University of 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.

Trials. 2019. 20 (Suppl 1):579 – P203.

This text box is where the unabridged thesis included the following third party copyrighted material:

Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, Gamble C. *Randomised trials involving surgery do not report considerations of learning and clustering effects*. Journal of Clinical Epidemiology. 2019; 107: 27 – 35.

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

This text box is where the unabridged thesis included the following third party copyrighted material:

Conroy EJ, Blazeby JM, Burnside G, Cook JA, Gamble C. *Design and analysis of the learning curve and clustering effects in randomised surgical trials*. Meeting abstracts from the 4th International Clinical Trial Methodology Conference (ICTMC 2017). Trials. 2017. 18 (Suppl 1):200 - P129..

doi: https://doi.org/10.1186/s13063-017-1902-y

This text box is where the unabridged thesis included the following third party copyrighted material:

Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, 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; 113: 28-35.

doi: https://doi.org/10.1016/j.jclinepi.2019.05.007

This text box is where the unabridged thesis included the following third party copyrighted material:

Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, Gamble C. *Managing learning and clustering effects at the design stage in randomised surgical trials: a retrospective cohort of trial funding applications.* Meeting abstracts from the 5th International Clinical Trial Methodology Conference (ICTMC 2019). Trials. 2019. 20 (Suppl 1):579 – P20.

doi: https://doi.org/10.1186/s13063-019-3688-6

This text box is where the unabridged thesis included the following third party copyrighted material:

Conroy EJ, Blazeby JM, Burnside G, Cook JA, Gamble C. *Managing clustering effects and learning effects in the design and analysis of multicentre randomised trials: a survey to establish current practice*. Trials. **21,** 433 (2020).

doi: https://doi.org/10.1186/s13063-020-04318-x

This text box is where the unabridged thesis included the following third party copyrighted material:

Conroy EJ, Blazeby JM, Burnside G, Cook JA, Gamble C. *Allowance for learning and clustering effects in the design and analysis of multicentre randomised trials: current practice and experiences.* Meeting abstracts from the 5th International Clinical Trial Methodology Conference (ICTMC 2019). Trials. 2019. 20 (Suppl 1):579 – P203.

doi: https://doi.org/10.1186/s13063-019-3688-6

Publications referred to in this thesis

The following publications are publications that I have co-authored and referred to in this thesis:

The BASICS trial

CL Mallucci, MD Jenkinson, EJ Conroy, JC Hartley, M Brown, J Dalton, T Kearns, T Moitt, MJ Griffiths, G Culeddu, T Solomon, D Hughes, C Gamble[4] Antibiotic or silver versus standard ventriculoperitoneal shunts (BASICS): a multicentre, single-blinded, randomised trial and economic evaluation. Lancet. 2019; 394: 1530-39.

CL Mallucci, MD Jenkinson, EJ Conroy, JC Hartley, M Brown, T Moitt, J Dalton, T Kearns, MJ Griffths, G Culeddu, T Solomon, D Hughes, Gamble C. Silver-impregnated, antibiotic-impregnated or non-impregnated ventriculoperitoneal shunts to prevent shunt infection: the BASICS three-arm RCT. Health Technol Assess 2020; 24 (17).

The TOPS trial

W Shaw, G Semb, A Lohmander, C Persson, E Willadsen, J Clayton-Smith, I Trindade, K Munro, C Gamble, N Harman, EJ Conroy, D Weichart, P Williamson. Timing Of Primary Surgery for cleft palate (TOPS): protocol for a randomised trial of palate surgery at 6 months versus 12 months of age. BMJ Open. 2019;9:e029780.

EJ Conroy, R Cooper, W Shaw, C Persson, E Willadsen, K Munro, PR Williamson, T Walsh, C Gamble, on behalf of the TOPS trial management group. A randomised controlled trial comparing palate surgery at 6 months versus 12 months of age (the TOPS trial): a statistical analysis plan. Trials. 2021 Jan 4; 22 (1): 5.

The NERVES trial

M Wilby, C Hopkins, E Bedson, G Burnside, B Conroy, D Hughes, M Sharma, S Clark and P Williamson. Nerve root block versus surgery (NERVES) for the treatment of radicular pain secondary to a prolapsed intervertebral disc herniation: study protocol for a multi-centre randomised controlled trial. Trials. 2018; 19: 475. DOI: 10.1186/s13063-018-2677-5.

Appendix 2 – Supplementary material from Chapter 2

Appendix Material 1: List of eligible guidance documents

The following guidance documents were identified using the EQUATOR Network search engine:

1. Bilbro NA, Hirst A, Paez A, Vasey B, Pufulete M, Sedrakyan A, McCulloch P; IDEAL Collaboration Reporting Guidelines Working Group. The IDEAL Reporting Guidelines: A Delphi Consensus Statement Stage specific recommendations for reporting the evaluation of surgical innovation. Ann Surg. 2020.
2. Elias KM, Stone AB, McGinigle K, Tankou JI, Scott MJ, Fawcett WJ, Demartines N, Lobo DN, Ljungqvist O, Urman RD; ERAS® Society and ERAS® USA. The Reporting on ERAS Compliance, Outcomes, and Elements Research (RECOvER) Checklist: A Joint Statement by the ERAS® and ERAS® USA Societies. World J Surg. 2019 Jan;43(1):1-8.
3. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P, Berlin J, Senn S, Day S, Barbachano Y, Loder E. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337-2343.
4. Jackson DL. Reporting results of latent growth modeling and multilevel modeling analyses: some recommendations for rehabilitation psychology. Rehabil Psychol. 2010;55(3):272-285.
5. de Jager DJ, de Mutsert R, Jager KJ, Zoccali C, Dekker FW. Reporting of interaction. Nephron Clin Pract. 2011;119(2):c158-161.
6. Kalil AC, Mattei J, Florescu DF, Sun J, Kalil RS. Recommendations for the assessment and reporting of multivariable logistic regression in transplantation literature. Am J Transplant. 2010;10(7):1686-1694.

7. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials. 2010;11:85.
8. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. Int J Nurs Stud. 2015;52(1):5-9.
9. van de Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. Guidelines for Reporting on Latent Trajectory Studies (GRoLTS). Structural Equation Modeling: A Multidisciplinary Journal. 2017;24(3):451-467.
10. Schreiber JB. Latent Class Analysis: An example for reporting results. Res Social Adm Pharm. 2016.
11. Vanhie A, Meuleman C, Tomassetti C, Timmerman D, D'Hoore A, Wolthuis A, Van Cleynenbreugel B, Dancet E, Van den Broeck U, Tsaltas J, Renner SP, Ebert AD, Carmona F, Abbott J, Stepniewska A, Taylor H, Saridogan E, Mueller M, Keckstein J, Pluchino N, Janik G, Zupi E, Minelli L, Cooper M, Dunselman G, Koh C, Abrao M, Chapron C, D'Hooghe T. Consensus on Recording Deep Endometriosis Surgery: the CORDES statement. Hum Reprod. 2016 Apr 19. pii: dew067.
12. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med. 2007;357(21):2189-2194.

The following guidance documents were identified by the targeted search:

Identified from UK funding bodies (n=2)

1. Clinical Trials Toolkit –Planning a Randomised Controlled Trial - Points to Consider: Funding proposal National Institute for Health Research web site: National Institute for Health Research; [Available from: https://www.ct-toolkit.ac.uk/routemap/trial-planning-and-design/downloads/planning-a-randomised-controlled-trial.pdf/.

2. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions Medical Research Council web site: Medical Research Council; 2019 [Available from: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/.

Identified from regulators (n=6)

3. Committee for Medicinal Products for Human Use (CHMP): Guideline on adjustment for baseline covariates in clinical trials European Medicines Agency Science Medicines Health web site: European Medicines Agency 2015 [Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf
4. Group IEW. ICH Harmonised Tripartite Guideline: Structure and Content of Clinical Study Reports E3 ICH Harmonisation for better health web site: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 1995 [Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf.
5. Group IEW. ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6(R1) ICH Harmonisation for better health web site: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 1996 [Available from: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf.
6. Group IEW. ICH Harmonised Tripartite Guideline: General Considerations for Clinical Trials E8 ICH Harmonisation for better health web site: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 1998 [Available from:

https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-8-general-considerations-clinical-trials-step-5_en.pdf

7. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. Stat Med. 1999;18(15):1905-42.
8. Medicines & Healthcare products Medicines Agency (MHRA): Guidance on legislation. Clinical investigations of medical devices – statistical considerations. Medicines & Healthcare products Regulatory Agency 2021. [Available from: Statistical_considerations_clinical_investigations_-_May_2021.pdf (publishing.service.gov.uk)

Identified from medical journals (n=8)

9. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Balliol Collaboration. Evaluation and stages of surgical innovations. Lancet. 2009;374(9695):1089-96.
10. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, DJ C, et al. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. Annals of Internal Medicine. American College of Physicians; 2017 Jul 4;167(1):40. PMID: 28630973
11. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P; CONSORT Group. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments. Ann Intern Med. 2008 Feb 19;148(4):W60-6. doi: 10.7326/0003-4819-148-4-200802190-00008-w1. PMID: 18283201.
12. Campbell MK, Piaggio G, Elbourne DR, Altman DG; for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012 Sep 4;345:e5661. PMID: 22951546

13. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, et al. Challenges in evaluating surgical innovation. Lancet. 2009;374(9695):1097-104.
14. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374(9695):1105-12.
15. Schulz K F, Altman D G, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials BMJ 2010; 340 :c332 doi:10.1136/bmj.c332
16. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D for the CONSORT and Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008; 337;a2390. PMID: 19001484

Appendix Table 1: Key criteria coverage across documents summary

Document	Key criteria [A]											Total
	Design						Analysis					
	D1	D2	D3	D4	D5	D6	A1	A2	A3	A4	A5	n
Barkun, 2009	No	No	Yes	Yes	No	No	No	No	No	No	No	2
McCulloch, 2009	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	6
Bilbro, 2021	No	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	5
Boutron, 2008	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	6
Boutron, 2017	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	8
Campbell, 2012	Yes	Yes	No	No	Yes	Yes	No	No	No	No	Yes	5
MRC, 2019	Yes	No	Yes	Yes	Yes	No	No	No	No	Yes	Yes	6
Elias, 2019	No	Yes	Yes	No	No	No	No	No	No	No	No	2
Ergina, 2009	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	4
EMA, 2015	No	No	No	No	No	Yes	Yes	Yes	No	No	Yes	4
ICH E3, 1995	No	No	Yes	No	No	Yes	No	No	Yes	No	No	3
ICH E6, 1996	No	Yes	No	No	No	No	No	No	No	No	No	1

Appendix Table 1: Key criteria coverage across documents summary

Document	Key criteria [A]											Total
	Design						Analysis					
	D1	D2	D3	D4	D5	D6	A1	A2	A3	A4	A5	n
ICH E8, 1998	No	No	Yes	Yes	No	No	No	No	No	No	No	2
ICH E9, 1999	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	8
Jackson, 2010	Yes	Yes	No	No	Yes	No	No	No	No	Yes	Yes	5
Jager, 2011	No	No	No	No	No	No	No	No	No	No	No	0
Gamble, 2017	No	No	No	No	No	No	No	No	No	No	No	0
Kalil, 2010	No	No	No	No	No	No	No	No	No	No	No	0
Kent, 2010	No	No	No	No	No	No	No	No	No	No	No	0
Lang, 2015	No	No	No	No	No	No	No	No	No	No	No	0
MHRA, 2021	Yes	Yes	No	No	Yes	Yes	No	No	No	No	No	4
NIHR Toolkit, 2016	No	Yes	Yes	No	Yes	No	No	No	No	No	No	3
Schoot, 2017	No	No	No	No	No	No	No	No	No	No	No	0
Schreiber, 2016	No	No	No	No	No	No	No	No	No	No	No	0

Appendix Table 1: Key criteria coverage across documents summary

Document	Key criteria [A]											Total
	Design						Analysis					
	D1	D2	D3	D4	D5	D6	A1	A2	A3	A4	A5	n
Schultz, 2011	Yes	No	Yes	No	No	No	No	No	No	No	No	2
MHRA, 2021	Yes	Yes	No	No	Yes	Yes	No	No	No	No	No	4
NIHR Toolkit, 2016	No	Yes	Yes	No	Yes	No	No	No	No	No	No	3
Vanhie et al, 2016	No	No	Yes	No	No	No	No	No	No	No	No	1
Wang, 2007	No	No	No	No	No	No	No	No	No	No	No	0
Zwarenstein, 2008	Yes	Yes	Yes	No	No	No	No	No	No	No	No	3
Total n (n/28%)	11 (39%)	13 (46%)	15 (54%)	8 (29%)	8 (29%)	6 (21%)	2 (7%)	2 (7%)	6 (21%)	4 (14%)	5 (18%)	

[A] **Key criteria:** D1: Choosing a trial design; D2: Considering who will deliver the intervention; D3: Ensuring that the intervention is standardised; D4: Anticipating changes in delivery over time; D5: Estimating the sample size; D6: Ensuring balance of treatment within centre and treatment provider; A1: When the randomization was stratified; A2: When analyzing the primary outcome; A3: Analysing multicenter trials; A4: Methods for investigating the learning curve; A5: Methods for investigating clustering.

Appendix 3 – Supplementary material from Chapter 3

Appendix Material 2: Search strategy

Phase 1: Identify surgical randomised trials

Step	Description	Results (n)
1	Search for surg* and rand* in Article Title, Abstract, Keywords Document type = Article No date range added as easy enough to filter by year	120, 441

Phase 2: Identify outputs from key journals

Step	Description	Results (n)
2	Search for Source Title "British Medical Journal" All document types No date range added	93, 041
3	Search for Source Title "BMJ" All document types No date range added	64, 620
4	Search for Source Title "Journal of the american medical association" All document types No date range added	188, 051
5	Search for Source Title "JAMA" All document types No date range added	78, 788
6	Search for Source Title "Lancet"	466, 129

	All document types No date range added	
7	Search for Source Title "New England Journal of Medicine" All document types No date range added	76, 877
8	Search for ISSN "1745-6215" which is the ISSN number for Trials All document types No date range added	2, 646
9	Search for ISSN "1366-5278" which is the ISSN number for HTA monograph series All document types No date range added	148

Phase 3: Identify outputs from leading surgical journals

List of the names of ten leading general surgical journals obtained (39). Details are as follows:

	Journal name	**ISSN number(s)**
1	Annals of Surgery	0003-4932 (print) & 1528-1140 (web)
2	British Journal of Surgery	0007-1323
3	Archives of Surgery - now JAMA Surgery	0004-0010 (archives of surgery before JAMA surgery) 2168-6254 (print) & 2168-6262 (web)
4	Surgery	0039-6060
5	American Journal of Surgery	0002-9610
6	Journal of American College of Surgeons	1072-7515
7	Current Problems in Surgery	0011-3840
8	American Surgeon	0003-1348

9	Australia and New Zealand Journal of Surgery	1445-2197 (web)
10	Surgery Today	0941-1291 (Print) & 1436-2813 (Online)

Step	Description	Results (n)
10	Search for ISSN "1528-1140" which is the ISSN number for Annals of Surgery (web) All document types No date range added	4, 021
11	Search for ISSN "0003-4932" which is the ISSN number for Annals of Surgery (print) All document types No date range added	17, 826
12	Search for ISSN "0007-1323" which is the ISSN number for British Journal of Surgery All document types No date range added	20, 660
13	Search for ISSN "2168-6254" which is the ISSN number for JAMA Surgery (print) All document types No date range added	856
14	Search for ISSN "2168-6262" which is the ISSN number for JAMA Surgery (web) All document types No date range added	19

15	Search for ISSN "0004-0010" which is the ISSN number for Annals of Surgery (before JAMA surgery) All document types No date range added	15, 777
16	Search for ISSN "0039-6060" which is the ISSN number for Surgery All document types No date range added	22, 115
17	Search for ISSN "0002-9610" which is the ISSN number for American Journal of Surgery All document types No date range added	30, 148
18	Search for ISSN "1072-7515" which is the ISSN number for Journal of American College of Surgeons All document types No date range added	6, 816
19	Search for ISSN "0011-3840" which is the ISSN number for Current Problems in Surgery All document types No date range added	1, 268
20	Search for ISSN "0003-1348" which is the ISSN number for American Surgeon All document types No date range added	13, 749
21	Search for ISSN "1445-2197" which is the ISSN number for Australia and New Zealand Journal of Surgery All document types No date range added	3, 672

22	Search for ISSN "0941-1291" which is the ISSN number for Surgery Today (print) All document types No date range added	5, 910
23	Search for ISSN "1436-2813" which is the ISSN number for Surgery Today (web) All document types No date range added	5, 103

Phase 4: Identify outputs from leading surgical journals

Step	Description	Results (n)
24	Merge general journals #4 OR #7 OR #6 OR #8 OR #9 OR #10 OR #12 OR #13	841, 424
25	Merge surgical journal (part 1) #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20	77, 393
26	Merge surgical journal (part 2) #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27	63, 079
27	Merge all journals #28 OR #29 OR #30	1, 021, 021
28	Obtain surgical RCTs from all journals #3 AND #31	6, 176

Appendix Material 3: Data extraction form

The following details were extracted from eligible articles:

SECTION 1: Article details

1.1. Journal (CATEGORICAL – *American Journal of Surgery / American Surgeon / Annals of Surgery / ANZ Journal of Surgery / BMJ (Online) / BMJ Open / BMJ Supportive and Palliative Care / British Journal of Surgery / Health Technology Assessment / JAMA – Journal of the American Medical Association / JAMA Dermatology / JAMA Facial Plastic Surgery / JAMA Opthalmology / JAMA Otalaryngology – Head and Neck Surgery / JAMA Surgery / Journal of the American College of Surgeons / Surgery (United States) / Surgery Today / The Lancet / The Lancet Diabetes and Endcrinology / The Lancet Neurology / The Lancet Oncology / The Lancet Respiratory Medicine / The New England Journal of Medicine / Trials*)

1.2. Year published (CATEGORICAL – *2014 / 2015 / 2016*)

1.3. Funder origin (CATEGORICAL – *by COUNTRY*)

SECTION 2: Trial rationale and design

2.1. Type of trial (CATEGORICAL – *Definitive / Pilot or feasibility)*

2.2. Trial design (CATEGORICAL – *Cluster / Crossover / Factorial / Parallel / N of 1)*

2.3. Number of trial arms (NUMERIC)

2.4. Blinding (BINARY – *Yes / No*)

2.5. Expertise based design (CATEGORICAL – *Pure, professionals delivering only one intervention / Hybrid, some professionals could deliver both)*

2.6. Allocation of treatment provider in expertise based design (FREETEXT)

SECTION 3: Intervention of interest

3.1. Intervention of interest (CATEGORICAL - *Surgery occurred but was not intervention of interest / Surgery occurred and was the intervention of interest*)

3.2. Comparator when surgery was intervention of interest (CATEGORICAL *Surgery / Medical / Other e.g. active monitoring*)

3.3. Surgical comparator in trials comparing two surgeries (CATEGORICAL - *Comparing different components of the same intervention / Different surgical interventions / Different time points of the same intervention*)

SECTION 4: Recruitment

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

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

SECTION 5: Centre or surgeon credentials

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

5.2. Centre credentials (FREETEXT)

5.3. Surgeon credentials (FREETEXT)

SECTION 6: Randomisation

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

6.2. If stratified, stratified by centre (BINARY – *Yes / No*) or surgeon (BINARY – *Yes / No*)

SECTION 7: Considerations of learning and/or clustering of included centres and/or surgeons

7.1. Descriptive (non-outcome) relating to centre reported e.g. number of patients per centre (FREETEXT)

7.2. Descriptive (non-outcome) relating to surgeon reported e.g. surgeon caseload (FREETEXT)

7.3. Outcomes (FREETEXT)

7.4. Analysis planned to address multiple centre effect adjustments (BINARY – *Yes / No*)

7.4a. If yes, approach used to check for effect (CATEGORICAL – *Term in regression model / Separate exploratory analysis / Other – free text*)

7.4b. If accounted for, how effect was specified (CATEGORICAL – *Fixed / Random / Time-varying / Unclear*)

7.5. Analysis planned to address multiple surgeon effect adjustments (BINARY – *Yes / No*)

7.5a. If yes, approach used to check for effect (CATEGORICAL – *Term in regression model / Separate exploratory analysis / Other – free text*)

7.5b. If accounted for, how effect was specified (CATEGORICAL – *Fixed / Random / Time-varying / Unclear*)

Appendix Material 4: List of eligible publications

1. Abdel-Wahab M, Mehilli J, Frerker C, Neumann FJ, Kurz T, Tölg R, et al. Comparison of balloon-expandable vs self-expandable valves in patients undergoing transcatheter aortic valve replacement: The CHOICE randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;311(15):1503-14.

2. Abo-Ryia MH, El-Khadrawy OH, Moussa GI, Saleh AM. Prospective randomized evaluation of open preperitoneal versus preaponeurotic primary elective mesh repair for paraumbilical hernias. Surgery Today. 2015;45(4):429-33.

3. Acker MA, Parides MK, Perrault LP, Moskowitz AJ, Gelijns AC, Voisine P, et al. Mitral-valve repair versus replacement for severe ischemic mitral regurgitation. New England Journal of Medicine. 2014;370(1):23-32.

4. Ackland GL, Iqbal S, Paredes LG, Toner A, Lyness C, Jenkins N, et al. Individualised oxygen delivery targeted haemodynamic therapy in high-risk surgical patients: A multicentre, randomised, double-blind, controlled, mechanistic trial. The Lancet Respiratory Medicine. 2015;3(1):33-41.

5. Adams DH, Popma JJ, Reardon MJ, Yakubov SJ, Coselli JS, Deeb GM, et al. Transcatheter aortic-valve replacement with a self-expanding prosthesis. New England Journal of Medicine. 2014;370(19):1790-8.

6. Aida T, Furukawa K, Suzuki D, Shimizu H, Yoshidome H, Ohtsuka M, et al. Preoperative immunonutrition decreases postoperative complications by modulating prostaglandin E2 production and T-cell differentiation in patients undergoing pancreatoduodenectomy. Surgery (United States). 2014;155(1):124-33.

7. Aird LNF, Bristol SG, Phang PT, Raval MJ, Brown CJ. Randomized double-blind trial comparing the cosmetic outcome of cutting diathermy versus scalpel for skin incisions. British Journal of Surgery. 2015;102(5):489-94.

8. Armañanzas L, Ruiz-Tovar J, Arroyo A, García-Peche P, Armañanzas E, Diez M, et al. Prophylactic mesh vs suture in the closure of the umbilical trocar site after laparoscopic

cholecystectomy in high-risk patients for incisional hernia. a randomized clinical trial. Journal of the American College of Surgeons. 2014;218(5):960-8.

9. Ausen K, Fossmark R, Spigset O, Pleym H. Randomized clinical trial of topical tranexamic acid after reduction mammoplasty. British Journal of Surgery. 2015;102(11):1348-53.

10. Ayoub N, Ghassemi A, Rana M, Gerressen M, Riediger D, Hölzle F, et al. Evaluation of computer-assisted mandibular reconstruction with vascularized iliac crest bone graft compared to conventional surgery: A randomized prospective clinical trial. Trials. 2014;15(1).

11. Bai X, Zhang Q, Gao S, Lou J, Li G, Zhang Y, et al. Duct-to-Mucosa vs Invagination for Pancreaticojejunostomy after Pancreaticoduodenectomy: A Prospective, Randomized Controlled Trial from a Single Surgeon. Journal of the American College of Surgeons. 2016;222(1):10-8.

12. Balta AZ, Ozdemir Y, Sucullu I, Ilker Filiz A, Yucel E, Akin ML. The effect of early warm plastic bag application on postoperative pain after hemorrhoidectomy: A prospective randomized controlled trial. American Surgeon. 2015;81(2):182-6.

13. Barber MD, Brubaker L, Burgio KL, Richter HE, Nygaard I, Weidner AC, et al. Comparison of 2 transvaginal surgical approaches and perioperative behavioral therapy for apical vaginal prolapse: The OPTIMAL randomized trial. JAMA - Journal of the American Medical Association. 2014;311(10):1023-34.

14. Barone MA, Widmer M, Arrowsmith S, Ruminjo J, Seuc A, Landry E, et al. Breakdown of simple female genital fistula repair after 7 day versus 14 day postoperative bladder catheterisation: A randomised, controlled, open-label, non-inferiority trial. The Lancet. 2015;386(9988):56-62.

15. Bath-Hextall F, Ozolins M, Armstrong SJ, Colver GB, Perkins W, Miller PSJ, et al. Surgical excision versus imiquimod 5% cream for nodular and superficial basal-cell carcinoma (SINS): A multicentre, non-inferiority, randomised controlled trial. The Lancet Oncology. 2014;15(1):96-105.

16. Bensaadi H, Paolino L, Valenti A, Polliand C, Barrat C, Champault G. Intraperitoneal tension-free repair of a small midline ventral abdominal wall hernia: Randomized study with a mean follow-up of 3 years. American Surgeon. 2014;80(1):57-65.

17. Berdah SV, Mariette C, Denet C, Panis Y, Laurent C, Cotte E, et al. A multicentre, randomised, controlled trial to assess the safety, ease of use, and reliability of hyaluronic acid/carboxymethylcellulose powder adhesion barrier versus no barrier in colorectal laparoscopic surgery. Trials. 2014;15(1).

18. Bhandari M, Jeray KJ, Petrisor BA, Devereaux PJ, Heels-Ansdell D, Schemitsch EH, et al. A Trial of Wound Irrigation in the Initial Management of Open Fracture Wounds. New England Journal of Medicine. 2015;373(27):2629-41.

19. Bill-Axelson A, Holmberg L, Garmo H, Rider JR, Taari K, Busch C, et al. Radical prostatectomy or watchful waiting in early prostate cancer. New England Journal of Medicine. 2014;370(10):932-42.

20. Bingener J, Skaran P, McConico A, Novotny P, Wettstein P, Sletten DM, et al. A double-blinded randomized trial to compare the effectiveness of minimally invasive procedures using patient-reported outcomes. Journal of the American College of Surgeons. 2015;221(1):111-21.

21. Boelens PG, Heesakkers FFBM, Luyer MDP, Van Barneveld KWY, De Hingh IHJT, Nieuwenhuijzen GAP, et al. Reduction of postoperative ileus by early enteral nutrition in patients undergoing major rectal surgery: Prospective, randomized, controlled trial. Annals of Surgery. 2014;259(4):649-55.

22. Boermeester MA. Effect of a ward-based pharmacy team on preventable adverse drug events in surgical patients (SUREPILL study). British Journal of Surgery. 2015;102(10):1204-12.

23. Bonjer HJ, Deijen CL, Abis GA, Cuesta MA, Van Der Pas MHGM, De Lange-De Klerk ESM, et al. A randomized trial of laparoscopic versus open surgery for rectal cancer. New England Journal of Medicine. 2015;372(14):1324-32.

24. Bonrath EM, Dedy NJ, Gordon LE, Grantcharov TP. Comprehensive surgical coaching enhances surgical skill in the operating room: A randomized controlled trial. Annals of Surgery. 2015;262(2):205-12.

25. Bove T, Zangrillo A, Guarracino F, Alvaro G, Persi B, Maglioni E, et al. Effect of fenoldopam on use of renal replacement therapy among patients with acute kidney injury after cardiac surgery: A randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;312(21):2244-53.

26. Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patientreported outcome measures-results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: A cluster randomised controlled study. BMJ Open. 2015;5(7).

27. Brittenden J, Cotton SC, Elders A, Ramsay CR, Norrie J, Burr J, et al. A randomized trial comparing treatments for varicose veins. New England Journal of Medicine. 2014;371(13):1218-27.

28. Brittenden J, Cotton SC, Elders A, Tassie E, Scotland G, Ramsay CR, et al. Clinical effectiveness and cost-effectiveness of foam sclerotherapy, endovenous laser ablation and surgery for varicose veins: Results from the comparison of LAser, Surgery and foam Sclerotherapy (CLASS) randomised controlled trial. Health Technology Assessment. 2015;19(27):1-341.

29. Bruix J, Takayama T, Mazzaferro V, Chau GY, Yang J, Kudo M, et al. Adjuvant sorafenib for hepatocellular carcinoma after resection or ablation (STORM): A phase 3, randomised, double-blind, placebo-controlled trial. The Lancet Oncology. 2015;16(13):1344-54.

30. Büller HR, Bethune C, Bhanot S, Gailani D, Monia BP, Raskob GE, et al. Factor XI antisense oligonucleotide for prevention of venous thrombosis. New England Journal of Medicine. 2015;372(3):232-40.

31. Busse JW, Bhandari M, Einhorn TA, Heckman JD, Leung KS, Schemitsch E, et al. Trial to re-evaluate ultrasound in the treatment of tibial fractures (TRUST): A multicenter randomized pilot study. Trials. 2014;15(1).

32. Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. New England Journal of Medicine. 2015;372(11):1009-18.

33. Campbell I, Cavanagh S, Creighton J, French R, Banerjee S, Kerr E, et al. To infiltrate or not? Acute effects of local anaesthetic in breast surgery. ANZ Journal of Surgery. 2015;85(5):353-7.

34. Carr AJ, Cooper CD, Campbell MK, Rees JL, Moser J, Beard DJ, et al. Clinical effectiveness and cost-effectiveness of open and arthroscopic rotator cuff repair [the UK rotator cuff surgery (UKUFF) randomised trial]. Health Technology Assessment. 2015;19(80):1-217.

35. Carter JT, Kaplan JA, Nguyen JN, Lin MYC, Rogers SJ, Harris HW. A prospective, randomized controlled trial of single-incision laparoscopic vs conventional 3-port laparoscopic appendectomy for treatment of acute appendicitis. Journal of the American College of Surgeons. 2014;218(5):950-9.

36. Chagpar AB, Killelea BK, Tsangaris TN, Butler M, Stavris K, Li F, et al. A randomized, controlled trial of cavity shave margins in breast cancer. New England Journal of Medicine. 2015;373(6):503-10.

37. Clark TJ, Middleton LJ, Cooper NA, Diwakar L, Denny E, Smith P, et al. A randomised controlled trial of outpatient versus inpatient polyp treatment (OPT) for abnormal uterine bleeding. Health Technology Assessment. 2015;19(61):1-194.

38. Clarke M, Hogan V, Buck D, Shen J, Powell C, Speed C, et al. An external pilot study to test the feasibility of a randomised controlled trial comparing eye muscle surgery against active monitoring for childhood intermittent exotropia [X(T)]. Health Technology Assessment. 2015;19(39):1-176.

39. Compter A, van der Worp HB, Schonewille WJ, Vos JA, Boiten J, Nederkoorn PJ, et al. Stenting versus medical treatment in patients with symptomatic vertebral artery stenosis: A randomised open-label phase 2 trial. The Lancet Neurology. 2015;14(6):606-14.

40. Correa-Gallego C, Tan KS, Arslan-Carlon V, Gonen M, Denis SC, Langdon-Embry L, et al. Goal-Directed Fluid Therapy Using Stroke Volume Variation for Resuscitation after Low Central Venous Pressure-Assisted Liver Resection: A Randomized Clinical Trial. Journal of the American College of Surgeons. 2015;221(2):591-601.

41. Costa ML, Achten J, Caroline P, Parsons NR, Rangan A, Tubeuf S, et al. UK DRAFFT: A randomised controlled triaof percutaneous fixation with kirschner wires versus volar locking-plate fixation in the treatment of adult patients with a dorsally displaced fracture of the distal radius. Health Technology Assessment. 2015;19(17).

42. Costa ML, Achten J, Parsons NR, Rangan A, Griffin D, Tubeuf S, et al. Percutaneous fixation with Kirschner wires versus volar locking plate fixation in adults with dorsally displaced fracture of distal radius: Randomised controlled trial. BMJ (Online). 2014;349.

43. Cotton PB, Durkalski V, Romagnuolo J, Pauls Q, Fogel E, Tarnasky P, et al. Effect of endoscopic sphincterotomy for suspected sphincter of oddi dysfunction on pain-related disability following cholecystectomy: The EPISOD randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;311(20):2101-9.

44. Courcoulas AP, Goodpaster BH, Eagleton JK, Belle SH, Kalarchian MA, Lang W, et al. Surgical vs medical treatments for type 2 diabetes mellitus: A randomized clinical trial. JAMA Surgery. 2014;149(7):707-15.

45. Custis T, Armstrong AW, King TH, Sharon VR, Eisen DB. Effect of adhesive strips and dermal sutures vs dermal sutures only on wound closure: A randomized clinical trial. JAMA Dermatology. 2015;151(8):862-7.

46. Czarnetzki C, Elia N, Frossard JL, Giostra E, Spahr L, Waeber JL, et al. Erythromycin for gastric emptying in patients undergoing general anesthesia for emergency surgery a randomized clinical trial. JAMA Surgery. 2015;150(8):730-7.

47. Da Costa DW, Bouwense SA, Schepers NJ, Besselink MG, Van Santvoort HC, Van Brunschot S, et al. Same-admission versus interval cholecystectomy for mild gallstone pancreatitis (PONCHO): A multicentre randomised controlled trial. The Lancet. 2015;386(10000):1261-8.

48. Daud WNW, Thompson SK, Jamieson GG, Devitt PG, Martin IJG, Watson DI. Randomized controlled trial of laparoscopic anterior 180° partial versus posterior 270° partial fundoplication. ANZ Journal of Surgery. 2015;85(9):668-72.

49. Davey C, Zoumot Z, Jordan S, McNulty WH, Carr DH, Hind MD, et al. Bronchoscopic lung volume reduction with endobronchial valves for patients with heterogeneous emphysema and intact interlobar fissures (the BeLieVeR-HIFi study): A randomised controlled trial. The Lancet. 2015;386(9998):1066-73.

50. Day AR, Smith RVP, Scott MJP, Fawcett WJ, Rockall TA. Randomized clinical trial investigating the stress response from two different methods of analgesia after laparoscopic colorectal surgery. British Journal of Surgery. 2015;102(12):1473-9.

51. Dayan SH, Ashourian N. Polydioxanone absorbable plate for cartilaginous grafting in endonasal rhinoplasty: A randomized clinical trial. JAMA Facial Plastic Surgery. 2016;18(1):47-53.

52. De Cruz P, Kamm MA, Hamilton AL, Ritchie KJ, Krejany EO, Gorelik A, et al. Crohn's disease management after intestinal resection: A randomised trial. The Lancet. 2015;385(9976):1406-17.

53. Deerenberg EB, Harlaar JJ, Steyerberg EW, Lont HE, Van Doorn HC, Heisterkamp J, et al. Small bites versus large bites for closure of abdominal midline incisions (STITCH): A double-blind, multicentre, randomised controlled trial. The Lancet. 2015;386(10000):1254-60.

54. Denost Q, Adam JP, Rullier A, Buscail E, Laurent C, Rullier E. Perineal transanal approach: A new standard for laparoscopic sphincter-saving resection in low rectal cancer, a randomized trial. Annals of Surgery. 2014;260(6):993-9.

55. Detollenaere RJ, Den Boon J, Stekelenburg J, IntHout J, Vierhout ME, Kluivers KB, et al. Sacrospinous hysteropexy versus vaginal hysterectomy with suspension of the uterosacral ligaments in women with uterine prolapse stage 2 or higher: Multicentre randomized non-inferiority trial. BMJ (Online). 2015;351.

56. Devereaux PJ, Mrkobrada M, Sessler DI, Leslie K, Alonso-Coello P, Kurz A, et al. Aspirin in patients undergoing noncardiac surgery. New England Journal of Medicine. 2014;370(16):1494-503.

57. Devereaux PJ, Sessler DI, Leslie K, Kurz A, Mrkobrada M, Alonso-Coello P, et al. Clonidine in patients undergoing noncardiac surgery. New England Journal of Medicine. 2014;370(16):1504-13.

58. Diener MK, Knebel P, Kieser M, Schüler P, Schiergens TS, Atanassov V, et al. Effectiveness of triclosan-coated PDS Plus versus uncoated PDS II sutures for prevention of surgical site infection after abdominal wall closure: The randomised controlled PROUD trial. The Lancet. 2014;384(9938):142-52.

59. Donker M, van Tienhoven G, Straver ME, Meijnen P, van de Velde CJH, Mansel RE, et al. Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (EORTC 10981-22023 AMAROS): A randomised, multicentre, open-label, phase 3 non-inferiority trial. The Lancet Oncology. 2014;15(12):1303-10.

60. Douketis JD, Spyropoulos AC, Kaatz S, Becker RC, Caprini JA, Dunn AS, et al. Perioperative bridging anticoagulation in patients with atrial fibrillation. New England Journal of Medicine. 2015;373(9):823-33.

61. Droeser RA, Dell-Kuster S, Kurmann A, Rosenthal R, Zuber M, Metzger J, et al. Long-term follow-up of a randomized controlled trial of Lichtenstein's operation versus mesh plug repair for inguinal hernia. Annals of Surgery. 2014;259(5):966-72.

62. Eid GM, McCloskey CA, Eagleton JK, Lee LB, Courcoulas AP. StomaphyX vs a sham procedure for revisional surgery toreduceregainedweightinroux-en-ygastric bypass patients a randomized clinical trial. JAMA Surgery. 2014;149(4):372-8.

63. El Messaoudi S, Nederlof R, Zuurbier CJ, van Swieten HA, Pickkers P, Noyez L, et al. Effect of metformin pretreatment on myocardial injury during coronary artery bypass surgery in patients without diabetes (MetCAB): A double-blind, randomised controlled trial. The Lancet Diabetes and Endocrinology. 2015;3(8):615-23.

64. Elamin G, Waters PS, Hamid H, O'Keeffe HM, Waldron RM, Duggan M, et al. Efficacy of a Laparoscopically Delivered Transversus Abdominis Plane Block Technique during Elective Laparoscopic Cholecystectomy: A Prospective, Double-Blind Randomized Trial. Journal of the American College of Surgeons. 2015;221(2):335-44.

65. Fakhry F, Spronk S, Van Der Laan L, Wever JJ, Teijink JAW, Hoffmann WH, et al. Endovascular revascularization and supervised exercise for peripheral artery disease and intermittent claudication: A randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;314(18):1936-44.

66. Fields AC, Gonzalez DO, Chin EH, Nguyen SQ, Zhang LP, Divino CM. Laparoscopic-Assisted Transversus Abdominis Plane Block for Postoperative Pain Control in Laparoscopic Ventral Hernia Repair: A Randomized Controlled Trial. Journal of the American College of Surgeons. 2015;221(2):462-9.

67. Fleshman J, Branda M, Sargent DJ, Boller AM, George V, Abbas M, et al. Effect of laparoscopic-assisted resection vs open resection of stage II or III rectal cancer on pathologic outcomes the ACOSOG Z6051 randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;314(13):1346-55.

68. Freedman SF, Lynn MJ, Beck AD, Bothun ED, Örge FH, Lambert SR. Glaucoma-related adverse events in the first 5 years after unilateral cataract removal in the infant aphakia treatment study. JAMA Ophthalmology. 2015;133(8):907-14.

69. Fujita J, Takiguchi S, Nishikawa K, Kimura Y, Imamura H, Tamura S, et al. Randomized controlled trial of the LigaSure vessel sealing system versus conventional open gastrectomy for gastric cancer. Surgery Today. 2014;44(9):1723-9.

70. Fukata S, Kawabata Y, Fujisiro K, Katagawa Y, Kuroiwa K, Akiyama H, et al. Haloperidol prophylaxis does not prevent postoperative delirium in elderly patients: a randomized, open-label prospective trial. Surgery today. 2014;44(12):2305-13.

71. Fyhn TJ, Knatten CK, Edwin B, Schistad O, Aabakken L, Kjosbakken H, et al. Randomized controlled trial of laparoscopic and open Nissen fundoplication in children. Annals of Surgery. 2015;261(6):1061-7.

72. Galizia G, Lieto E, De Vita F, Castellano P, Ferraraccio F, Zamboli A, et al. Modified versus standard D2 lymphadenectomy in total gastrectomy for nonjunctional gastric carcinoma with lymph node metastasis. Surgery (United States). 2015;157(2):285-96.

73. Gall TMH, Jacob J, Frampton AE, Krell J, Kyriakides C, Castellano L, et al. Reduced dissemination of circulating tumor cells with no-touch isolation surgical technique in patients with pancreatic cancer. JAMA Surgery. 2014;149(5):482-5.

74. García-Ureña MÁ, López-Monclús J, Hernando LA, Montes DM, Valle de Lersundi AR, Pavón CC, et al. Randomized controlled trial of the use of a large-pore polypropylene mesh to prevent incisional hernia in colorectal surgery. Annals of surgery. 2015;261(5):876-81.

75. Genser L, Trésallet C, Godiris-Petit G, Li Sun Fui S, Salepcioglu H, Royer C, et al. Randomized controlled trial of alfacalcidol supplementation for the reduction of hypocalcemia after total thyroidectomy. American Journal of Surgery. 2014;207(1):39-45.

76. Giles YS, Sarici IS, Tunca F, Sormaz IC, Salmaslioɨlu A, Adalet I, et al. The rate of operative success achieved with radioguided occult lesion localization and intraoperative ultrasonography in patients with recurrent papillary thyroid cancer. Surgery (United States). 2014;156(5):1116-26.

77. Gillinov AM, Gelijns AC, Parides MK, DeRose JJ, Jr., Moskowitz AJ, Voisine P, et al. Surgical ablation of atrial fibrillation during mitral-valve surgery. New England Journal of Medicine. 2015;372(15):1399-409.

78. Goldstein JN, Refaai MA, Milling TJ, Lewis B, Goldberg-Alberts R, Hug BA, et al. Four-factor prothrombin complex concentrate versus plasma for rapid vitamin K antagonist

reversal in patients needing urgent surgical or invasive interventions: A phase 3b, open-label, non-inferiority, randomised trial. The Lancet. 2015;385(9982):2077-87.

79. Gonenc M, Dural AC, Celik F, Akarsu C, Kocatas A, Kalayci MU, et al. Enhanced postoperative recovery pathways in emergency surgery: A randomised controlled clinical trial. American Journal of Surgery. 2014;207(6):807-14.

80. Gordin EA, Luginbuhl AL, Ortlip T, Heffelfinger RN, Krein H. Subcutaneous vs Intramuscular Botulinum Toxin split-face randomized study. JAMA Facial Plastic Surgery. 2014;16(3):193-8.

81. Gough AE, Chang S, Reddy S, Ferrigno L, Zerey M, Grotts J, et al. Periprosthetic anesthetic for postoperative pain after laparoscopic ventral hernia repair a randomized clinical trial. JAMA Surgery. 2015;150(9):835-40.

82. Gregson BA, Rowan EN, Francis R, McNamee P, Boyers D, Mitchell P, et al. Surgical trial in traumatic intracerebral haemorrhage (STITCH): A randomised controlled trial of early surgery compared with Initial conservative treatment. Health Technology Assessment. 2015;19(70):1-138.

83. Griffin D, Parsons N, Shaw E, Kulikov Y, Hutchinson C, Thorogood M, et al. Operative versus non-operative treatment for closed, displaced, intra-articular fractures of the calcaneus: Randomised controlled trial. BMJ (Online). 2014;349.

84. Halperin F, Ding SA, Simonson DC, Panosian J, Goebel-Fabbri A, Wewalka M, et al. Roux-en-Y gastric bypass surgery or lifestyle with intensive medical management in patients with type 2 diabetes: Feasibility and 1-year results of a randomized clinical trial. JAMA Surgery. 2014;149(7):716-26.

85. Hamill JK, Liley A, Hill AG. Rectus sheath block for laparoscopic appendicectomy: A randomized clinical trial. ANZ Journal of Surgery. 2015;85(12):951-6.

86. Handoll H, Brealey S, Rangan A, Keding A, Corbacho B, Jefferson L, et al. The ProFHer (PROximal fracture of the humerus: Evaluation by randomisation) trial – A pragmatic multicentre randomized controlled trial evaluating the clinical effectiveness and

cost-effectiveness of surgical compared with non-surgical treatment for proximal fracture of the humerus in adults. Health Technology Assessment. 2015;19(24):1-279.

87. Hausenloy DJ, Candilio L, Evans R, Ariti C, Jenkins DP, Kolvekar S, et al. Remote ischemic preconditioning and outcomes of cardiac surgery. New England Journal of Medicine. 2015;373(15):1408-17.

88. He L, Manche EE. Contralateral eye-to-eye comparison of wavefront-guided and wavefront-optimized photorefractive keratectomy: A randomized clinical trial. JAMA Ophthalmology. 2015;133(1):51-9.

89. Henderson MA, Burmeister BH, Ainslie J, Fisher R, Di Iulio J, Smithers BM, et al. Adjuvant lymph-node field radiotherapy versus observation only in patients with melanoma at high risk of further lymph-node field relapse after lymphadenectomy (ANZMTG 01.02/TROG 02.01): 6-year follow-up of a phase 3, randomised controlled trial. The Lancet Oncology. 2015;16(9):1049-60.

90. Hilton P, Armstrong N, Brennand C, Howel D, Shen J, Bryant A, et al. INVESTIGATE-I (INVasive Evaluation before Surgical Treatment of Incontinence Gives Added Therapeutic Effect?): a mixed-methods study to assess the feasibility of a future randomised controlled trial of invasive urodynamic testing prior to surgery for stress urinary incontinence in women. Health Technology Assessment. 2015;19(15):1-274.

91. Hilton P, Armstrong N, Brennand C, Howel D, Shen J, Bryant A, et al. A mixed methods study to assess the feasibility of a randomised controlled trial of invasive urodynamic testing versus clinical assessment and non-invasive tests prior to surgery for stress urinary incontinence in women: The INVESTIGATE-I study. Trials. 2015;16(1).

92. Hirao M, Kurokawa Y, Fujita J, Imamura H, Fujiwara Y, Kimura Y, et al. Long-term outcomes after prophylactic bursectomy in patients with resectable gastric cancer: Final analysis of a multicenter randomized controlled trial. Surgery (United States). 2015;157(6):1099-105.

93. Honda M, Hiki N, Nunobe S, Ohashi M, Kumagai K, Hashimoto Y, et al. Preoperative vs Postoperative Eradication of Helicobacter pylori in 150 Patients with Gastric Cancer: A

Randomized Controlled Trial. Journal of the American College of Surgeons. 2015;221(2):273-9.

94. Hong SJ, Kim BK, Shin DH, Nam CM, Kim JS, Ko YG, et al. Effect of intravascular ultrasound-guided vs angiography- guided everolimus-eluting stent implantation: The IVUS-XPL randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;314(20):2155-63.

95. Hong YS, Nam BH, Kim KP, Kim JE, Park SJ, Park YS, et al. Oxaliplatin, fluorouracil, and leucovorin versus fluorouracil and leucovorin as adjuvant chemotherapy for locally advanced rectal cancer after preoperative chemoradiotherapy (ADORE): An open-label, multicentre, phase 2, randomised controlled trial. The Lancet Oncology. 2014;15(11):1245-53.

96. Hu ZC, Chen D, Guo D, Liang YY, Zhang J, Zhu JY, et al. Randomized clinical trial of autologous skin cell suspension combined with skin grafting for chronic wounds. British Journal of Surgery. 2015;102(2):e117-e23 and e80-e81.

97. Huang G, Lau WY, Wang ZG, Pan ZY, Yuan SX, Shen F, et al. Antiviral therapy improves postoperative survival in patients with hepatocellular carcinoma. Annals of Surgery. 2015;261(1):56-66.

98. Hübner M, Blanc C, Roulin D, Winiker M, Gander S, Demartines N. Randomized clinical trial on epidural versus patient-controlled analgesia for laparoscopic colorectal surgery within an enhanced recovery pathway. Annals of surgery. 2015;261(4):648-53.

99. Iacobone M, Citton M, Zanella S, Scarpa M, Pagura G, Tropea S, et al. The effects of acupuncture after thyroid surgery: A randomized, controlled trial. Surgery (United States). 2014;156(6):1605-12.

100. Ikramuddin S, Blackstone RP, Brancatisano A, Toouli J, Shah SN, Wolfe BM, et al. Effect of reversible intermittent intra-abdominal vagal nerve blockade on morbid obesity: The ReCharge randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;312(9):915-22.

101. Imazio M, Brucato A, Ferrazzi P, Pullara A, Adler Y, Barosi A, et al. Colchicine for prevention of postpericardiotomy syndrome and postoperative atrial fibrillation: The COPPS-2 randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;312(10):1016-23.

102. Iranmanesh P, Frossard JL, Mugnier-Konrad B, Morel P, Majno P, Nguyen-Tang T, et al. Initial cholecystectomy vs sequential common duct endoscopic assessment and subsequent cholecystectomy for suspected gallstone migration: A randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;312(2):137-44.

103. Jang JY, Kang MJ, Heo JS, Choi SH, Choi DW, Park SJ, et al. A prospective randomized controlled study comparing outcomes of standard resection and extended resection, including dissection of the nerve plexus and various lymph nodes, in patients with pancreatic head cancer. Annals of Surgery. 2014;259(4):656-64.

104. Jeong SY, Park JW, Nam BH, Kim S, Kang SB, Lim SB, et al. Open versus laparoscopic surgery for mid-rectal or low-rectal cancer after neoadjuvant chemoradiotherapy (COREAN trial): Survival outcomes of an open-label, non-inferiority, randomised controlled trial. The Lancet Oncology. 2014;15(7):767-74.

105. Jeroukhimov I, Wiser I, Karasic E, Nesterenko V, Poluksht N, Lavy R, et al. Reduced postoperative chronic pain after tension-free inguinal hernia repair using absorbable sutures: A single-blind randomized clinical trial. Journal of the American College of Surgeons. 2014;218(1):102-7.

106. Jones KA, LaFerriere KA. Intravenous anesthesia with Bispectral Index Monitoring vs Inhalational Anesthesia for rhytidoplasty a randomized clinical trial. JAMA Facial Plastic Surgery. 2015;17(4):239-44.

107. Joo J, Custis T, Armstrong AW, King TH, Omlin K, Kappel ST, et al. Purse-string suture vs second intention healing: Results of a randomized, blind clinical trial. JAMA Dermatology. 2015;151(3):265-70.

108. Jørgensen LN, Rosenberg J, Al-Tayar H, Assaadzadeh S, Helgstrand F, Bisgaard T. Randomized clinical trial of single- versus multi-incision laparoscopic cholecystectomy. British Journal of Surgery. 2014;101(4):347-55.

109. Jovin TG, Chamorro A, Cobo E, De Miquel MA, Molina CA, Rovira A, et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. New England Journal of Medicine. 2015;372(24):2296-306.

110. Jüttler E, Unterberg A, Woitzik J, Bösel J, Amiri H, Sakowitz OW, et al. Hemicraniectomy in older patients with extensive middle-cerebral-artery stroke. New England Journal of Medicine. 2014;370(12):1091-100.

111. Kabul Gürbulak E, Gürbulak B, Akgün IE, Özel A, Akan D, Ömerołlu S, et al. Effects of totally extraperitoneal (TEP) and Lichtenstein hernia repair on testicular blood flow and volume. Surgery (United States). 2015;158(5):1297-303.

112. Kakkar AK, Agnelli G, Fisher W, George D, Lassen MR, Mismetti P, et al. Preoperative enoxaparin versus postoperative semuloparin thromboprophylaxis in major abdominal surgery: A randomized controlled trial. Annals of Surgery. 2014;259(6):1073-9.

113. Kapma MR, Dijksman LM, Reimerink JJ, De Groof AJ, Zeebregts CJ, Wisselink W, et al. Cost-effectiveness and cost-utility of endovascular versus open repair of ruptured abdominal aortic aneurysm in the Amsterdam acute aneurysm trial. British Journal of Surgery. 2014;101(3):208-15.

114. Kazui H, Miyajima M, Mori E, Ishikawa M, Hirai O, Kuwana N, et al. Lumboperitoneal shunt surgery for idiopathic normal pressure hydrocephalus (SINPHONI-2): An open-label randomised trial. The Lancet Neurology. 2015;14(6):585-94.

115. Kefalopoulou Z, Zrinzo L, Jahanshahi M, Candelario J, Milabo C, Beigi M, et al. Bilateral globus pallidus stimulation for severe Tourette's syndrome: A double-blind, randomised crossover trial. The Lancet Neurology. 2015;14(6):595-605.

116. Kehoe S, Hook J, Nankivell M, Jayson GC, Kitchener H, Lopes T, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer

(CHORUS): An open-label, randomised, controlled, non-inferiority trial. The Lancet. 2015;386(9990):249-57.

117. Kim HO, Lee SR, Choi WJ, Kim H. Early oral feeding following laparoscopic colorectal cancer surgery. ANZ Journal of Surgery. 2014;84(7-8):539-44.

118. Kirkpatrick AW, Roberts DJ, Faris PD, Ball CG, Kubes P, Tiruta C, et al. Active negative pressure peritoneal therapy after abbreviated laparotomy: The intraperitoneal vacuum randomized controlled trial. Annals of Surgery. 2015;262(1):38-46.

119. Klingberg-Allvin M, Cleeve A, Atuhairwe S, Tumwesigye NM, Faxelid E, Byamugisha J, et al. Comparison of treatment of incomplete abortion with misoprostol by physicians and midwives at district level in Uganda: A randomised controlled equivalence trial. The Lancet. 2015;385(9985):2392-8.

120. Kobayashi S, Ito M, Yamamoto S, Kinugasa Y, Kotake M, Saida Y, et al. Randomized clinical trial of skin closure by subcuticular suture or skin stapling after elective colorectal cancer surgery. British Journal of Surgery. 2015;102(5):495-500.

121. Kobayashi S, Takeda Y, Nakahira S, Tsujie M, Shimizu J, Miyamoto A, et al. Fibrin Sealant with Polyglycolic Acid Felt vs Fibrinogen-Based Collagen Fleece at the Liver Cut Surface for Prevention of Postoperative Bile Leakage and Hemorrhage: A Prospective, Randomized, Controlled Study. Journal of the American College of Surgeons. 2016;222(1):59-64.

122. Koehestanie P, De Jonge C, Berends FJ, Janssen IM, Bouvy ND, Greve JWM. The effect of the endoscopic duodenal-jejunal bypass liner on obesity and type 2 diabetes mellitus, a multicenter randomized controlled trial. Annals of Surgery. 2014;260(6):984-92.

123. Kubota A, Goda T, Tsuru T, Yonekura T, Yagi M, Kawahara H, et al. Efficacy and safety of strong acid electrolyzed water for peritoneal lavage to prevent surgical site infection in patients with perforated appendicitis. Surgery Today. 2015;45(7):876-9.

124. Kunkler IH, Williams LJ, Jack WJL, Cameron DA, Dixon JM. Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): A randomised controlled trial. The Lancet Oncology. 2015;16(3):266-73.

125. Kurmann A, Fischer H, Dell-Kuster S, Rosenthal R, Audigé L, Schüpfer G, et al. Effect of intraoperative infiltration with local anesthesia on the development of chronic pain after inguinal hernia repair: A randomized, triple-blinded, placebo-controlled trial. Surgery (United States). 2015;157(1):144-54.

126. Lairmore TC, Govednik CM, Quinn CE, Sigmond BR, Lee CY, Jupiter DC. A randomized, prospective trial of operative treatments for hyperparathyroidism in patients with multiple endocrine neoplasia type 1. Surgery (United States). 2014;156(6):1326-34.

127. Lambert SR, Lynn MJ, Hartmann EE, DuBois L, Drews-Botsch C, Freedman SF, et al. Comparison of contact lens and intraocular lens correction of monocular aphakia during infancy: A randomized clinical trial of HOTV optotype acuity at age 4.5 years and clinical findings at age 5 years. JAMA Ophthalmology. 2014;132(6):676-82.

128. Lane TR, Kelleher D, Shepherd AC, Franklin IJ, Davies AH. Ambulatory varicosity avulsion later or synchronized (AVULS): a randomized clinical trial. Annals of surgery. 2015;261(4):654-61.

129. Li B, Xiang J, Zhang Y, Li H, Zhang J, Sun Y, et al. Comparison of Ivor-Lewis vs Sweet esophagectomy for esophageal squamous cell carcinoma a randomized clinical trial. JAMA Surgery. 2015;150(4):292-8.

130. Liang X, Han-Xin Z, Chang-E J, Chao W, Juan H, Guo-Quan G. Subcutaneous suture can accelerate wound healing of lower midline incision: A randomized controlled trial. American Surgeon. 2015;81(1):23-30.

131. Lobo MD, Sobotka PA, Stanton A, Cockcroft JR, Sulke N, Dolan E, et al. Central arteriovenous anastomosis for the treatment of patients with uncontrolled hypertension (the ROX CONTROL HTN study): A randomised controlled trial. The Lancet. 2015;385(9978):1634-41.

132. Löfgren J, Nordin P, Ibingira C, Matovu A, Galiwango E, Wladis A. A Randomized Trial of Low-Cost Mesh in Groin Hernia Repair. New England Journal of Medicine. 2016;374(2):146-53.

133. Long KL, Hamilton DA, Davenport DL, Bernard AC, Kearney PA, Chang PK. A prospective, controlled evaluation of the abdominal reapproximation anchor abdominal wall closure system in combination with VAC therapy compared with VAC alone in the management of an open abdomen. American Surgeon. 2014;80(6):567-71.

134. Lurje G, Raptis DA, Steinemann DC, Amygdalos I, Kambakamba P, Petrowsky H, et al. Cosmesis and body image in patients undergoing single-port versus conventional laparoscopic cholecystectomy: A multicenter double-blinded randomized controlled trial (SPOCC-trial). Annals of Surgery. 2015;262(5):728-35.

135. Macrae D, Grieve R, Allen E, Sadique Z, Morris K, Pappachan J, et al. A randomized trial of hyperglycemic control in pediatric intensive care. New England Journal of Medicine. 2014;370(2):107-18.

136. Maeda K, Hanai T, Sato H, Masumori K, Koide Y, Matsuoka H, et al. Intraluminal exfoliated cancer cells and effectiveness of bowel ligatures during sigmoidectomy for sigmoid colon cancer. Surgery Today. 2014;44(2):297-301.

137. Mahmoodzadeh H, Shoar S, Sirati F, Khorgami Z. Early initiation of oral feeding following upper gastrointestinal tumor surgery: a randomized controlled trial. Surgery Today. 2014;45(2):203-8.

138. Malskat WSJ, Giang J, De Maeseneer MGR, Nijsten TEC, Van Den Bos RR. Randomized clinical trial of 940- versus 1470-nm endovenous laser ablation for great saphenous vein incompetence. British Journal of Surgery. 2016;103(3):192-8.

139. Matsutani N, Dejima H, Takahashi Y, Kawamura M. Pregabalin reduces post-surgical pain after thoracotomy: a prospective, randomized, controlled trial. Surgery Today. 2015;45(11):1411-6.

140. Maurice-Szamburski A, Auquier P, Viarre-Oreal V, Cuvillon P, Carles M, Ripart J, et al. Effect of sedative premedication on patient experience after general anesthesia: A randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;313(9):916-25.

141. Mazaki T, Mado K, Masuda H, Shiono M, Tochikura N, Kaburagi M. A randomized trial of antibiotic prophylaxis for the prevention of surgical site infection after open mesh-plug hernia repair. American Journal of Surgery. 2014;207(4):476-84.

142. Mehanna R, Hennessy A, Mannion S, O'Leary G, Sheahan P. Effect of endotracheal tube size on vocal outcomes after thyroidectomy: A randomized clinical trial. JAMA Otolaryngology - Head and Neck Surgery. 2015;141(8):690-5.

143. Melissa CS, Bun TAY, Wing CK, Chung TY, Wai NEK, Tat LH. Randomized double-blinded prospective trial of fibrin sealant spray versus mechanical stapling in laparoscopic total extraperitoneal hernioplasty. Annals of Surgery. 2014;259(3):432-7.

144. Mentula P, Sammalkorpi H, Leppäniemi A. Laparoscopic surgery or conservative treatment for appendiceal abscess in adults? A randomized controlled trial. Annals of Surgery. 2015;262(2):237-42.

145. Meybohm P, Bein B, Brosteanu O, Cremer J, Gruenewald M, Stoppe C, et al. A multicenter trial of remote ischemic preconditioning for heart surgery. New England Journal of Medicine. 2015;373(15):1397-407.

146. Milone M, Musella M, Maietta P, Bianco P, Taffuri C, Salvatore G, et al. Intradermal absorbable sutures to close pilonidal sinus wounds: A safe closure method? Surgery Today. 2014;44(9):1638-42.

147. Mokini Z, Vitale G, Aletti G, Sacchi V, Mauri T, Colombo V, et al. Pain control with ultrasound-guided inguinal field block compared with spinal anesthesia after hernia surgery: A randomized trial. Surgery (United States). 2015;157(2):304-11.

148. Mol F, Van Mello NM, Strandell A, Strandell K, Jurkovic D, Ross J, et al. Salpingotomy versus salpingectomy in women with tubal pregnancy (ESEP study): An open-label, multicentre, randomised controlled trial. The Lancet. 2014;383(9927):1483-9.

149. Moulton CA, Gu CS, Law CH, Tandan VR, Hart R, Quan D, et al. Effect of PET before liver resection on surgical management for colorectal adenocarcinoma metastases a randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;311(18):1863-9.

150. Mouton R, Pollock J, Soar J, Mitchell DC, Rogers CA. Remote ischaemic preconditioning versus sham procedure for abdominal aortic aneurysm repair: An external feasibility randomized controlled trial. Trials. 2015;16(1).

151. Müller-Stich BP, Achtstätter V, Diener MK, Gondan M, Warschkow R, Marra F, et al. Repair of Paraesophageal Hiatal Hernias-Is a Fundoplication Needed? A Randomized Controlled Pilot Trial. Journal of the American College of Surgeons. 2015;221(2):602-10.

152. Müller-Stich BP, Linke GR, Senft J, Achtstätter V, Müller PC, Diener MK, et al. Laparoscopic mesh-augmented hiatoplasty with cardiophrenicopexy versus laparoscopic nissen fundoplication for the treatment of gastroesophageal reflux disease. Annals of Surgery. 2015;262(5):721-7.

153. Murphy GJ, Pike K, Rogers CA, Wordsworth S, Stokes EA, Angelini GD, et al. Liberal or restrictive transfusion after cardiac surgery. New England Journal of Medicine. 2015;372(11):997-1008.

154. Murray DW, MacLennan GS, Breeman S, Dakin HA, Johnston L, Campbell MK, et al. A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: The Knee Arthroplasty Trial (KAT). Health Technology Assessment. 2014;18(19):1-235.

155. Myles PS, Leslie K, Chan MTV, Forbes A, Peyton PJ, Paech MJ, et al. The safety of addition of nitrous oxide to general anaesthesia in at-risk patients having major non-cardiac surgery (ENIGMA-II): A randomised, single-blind trial. The Lancet. 2014;384(9952):1446-54.

156. Narayanaswamy A, Leung CK, Istiantoro DV, Perera SA, Ho C, Nongpiur ME, et al. Efficacy of selective laser trabeculoplasty in primary angle-closure glaucoma: A randomized clinical trial. JAMA Ophthalmology. 2015;133(2):206-12.

157. Ohira G, Kawahira H, Miyauchi H, Suzuki K, Nishimori T, Hanari N, et al. Synthetic polyglycomer short-term absorbable sutures vs. polydioxanone long-term absorbable sutures for preventing incisional hernia and wound dehiscence after abdominal wall closure: a

comparative randomized study of patients treated for gastric or colon cancer. Surgery Today. 2015;45(7):841-5.

158. Orlando G, Manzia TM, Sorge R, Iaria G, Angelico R, Sforza D, et al. One-shot versus multidose perioperative antibiotic prophylaxis after kidney transplantation: A randomized, controlled clinical trial. Surgery (United States). 2015;157(1):104-10.

159. Palter VN, Grantcharov TP. Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: A randomized controlled trial. Annals of Surgery. 2014;259(3):443-8.

160. Park SJ, Ahn JM, Kim YH, Park DW, Yun SC, Lee JY, et al. Trial of everolimus-eluting stents or bypass surgery for coronary disease. New England Journal of Medicine. 2015;372(13):1204-12.

161. Park SY, Choi GS, Park JS, Kim HJ, Park JA, Choi JI. Efficacy and safety of udenafil for the treatment of erectile dysfunction after total mesorectal excision of rectal cancer: A randomized, double-blind, placebo-controlled trial. Surgery (United States). 2015;157(1):64-71.

162. Pearse RM, Harrison DA, MacDonald N, Gillies MA, Blunt M, Ackland G, et al. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery a randomized clinical trial and systematic review. JAMA - Journal of the American Medical Association. 2014;311(21):2181-90.

163. Pless M, Stupp R, Ris HB, Stahel RA, Weder W, Thierstein S, et al. Induction chemoradiation in stage IIIA/N2 non-small-cell lung cancer: A phase 3 randomised trial. The Lancet. 2015;386(9998):1049-56.

164. Poortmans PM, Collette S, Kirkove C, Van Limbergen E, Budach V, Struikmans H, et al. Internal mammary and medial supraclavicular irradiation in breast cancer. New England Journal of Medicine. 2015;373(4):317-27.

165. Porter M, Shadbolt B, Stuart R. Primary ankle ligament augmentation versus modified brostrom-gould procedure: A 2-year randomized controlled trial. ANZ Journal of Surgery. 2015;85(1-2):44-8.

166. Porter MD, Shadbolt B. Randomized controlled trial of accelerated rehabilitation versus standard protocol following surgical repair of ruptured Achilles tendon. ANZ Journal of Surgery. 2015;85(5):373-7.

167. Prather JL, Tummel EK, Patel AB, Smith DJ, Gould LJ. Prospective Randomized Controlled Trial Comparing the Effects of Noncontact Low-Frequency Ultrasound with Standard Care in Healing Split-Thickness Donor Sites. Journal of the American College of Surgeons. 2015;221(2):309-18.

168. Primrose J, Falk S, Finch-Jones M, Valle J, O'Reilly D, Siriwardena A, et al. Systemic chemotherapy with or without cetuximab in patients with resectable colorectal liver metastasis: The New EPOC randomised controlled trial. The Lancet Oncology. 2014;15(6):601-11.

169. Primrose JN, Perera R, Gray A, Rose P, Fuller A, Corkhill A, et al. Effect of 3 to 5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: The FACS randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;311(3):263-70.

170. Pucher PH, Aggarwal R, Qurashi M, Singh P, Darzi A. Randomized clinical trial of the impact of surgical ward-care checklists on postoperative care in a simulated environment. British Journal of Surgery. 2014;101(13):1666-73.

171. Qvamme G, Axelsson CK, Lanng C, Mortensen M, Wegeberg B, Okholm M, et al. Randomized clinical trial of prevention of seroma formation after mastectomy by local methylprednisolone injection. British Journal of Surgery. 2015;102(10):1195-203.

172. Rahbari NN, Elbers H, Koch M, Vogler P, Striebel F, Bruckner T, et al. Randomized clinical trial of stapler versus clamp-crushing transection in elective liver resection. British Journal of Surgery. 2014;101(3):200-7.

173. Rammohan A, Sathyanesan J, Rajendran K, Pitchaimuthu A, Perumal SK, Balaraman K, et al. Synbiotics in surgery for chronic pancreatitis: Are they truly effective?: A single-blind prospective randomized control trial. Annals of Surgery. 2015;262(1):31-7.

174. Rangan A, Handoll H, Brealey S, Jefferson L, Keding A, Martin BC, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus the PROFHER randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;313(10):1037-47.

175. Rasmussen KC, Johansson PI, Højskov M, Kridina I, Kistorp T, Thind P, et al. Hydroxyethyl starch reduces coagulation competence and increases blood loss during major surgery: Results from a randomized controlled trial. Annals of Surgery. 2014;259(2):249-54.

176. Räty S, Pulkkinen J, Nordback I, Sand J, Victorzon M, Grönroos J, et al. Can laparoscopic cholecystectomy prevent recurrent idiopathic acute pancreatitis? Annals of Surgery. 2015;262(5):736-41.

177. Regimbeau JM, Fuks D, Pautrat K, Mauvais F, Haccart V, Msika S, et al. Effect of postoperative antibiotic administration on postoperative infection following cholecystectomy for acute calculous cholecystitis: A randomized clinical trial. JAMA - Journal of the American Medical Association. 2014;312(2):145-54.

178. Rintoul RC, Ritchie AJ, Edwards JG, Waller DA, Coonar AS, Bennett M, et al. Efficacy and cost of video-assisted thoracoscopic partial pleurectomy versus talc pleurodesis in patients with malignant pleural mesothelioma (MesoVATS): An open-label, randomised, controlled trial. The Lancet. 2014;384(9948):1118-27.

179. Robinson TN, Jones EL, Dunn CL, Dunne B, Johnson E, Townsend NT, et al. Separating the laparoscopic camera cord from the monopolar "Bovie" cord reduces unintended thermal injury from antenna coupling: A randomized controlled trial. Annals of Surgery. 2015;261(6):1056-60.

180. Rödel C, Graeven U, Fietkau R, Hohenberger W, Hothorn T, Arnold D, et al. Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy of locally advanced rectal cancer (the German CAO/ARO/AIO-04 study): Final results of the multicentre, open-label, randomised, phase 3 trial. The Lancet Oncology. 2015;16(8):979-89.

181. Rodríguez A, Taurà P, García Domingo MI, Herrero E, Camps J, Forcada P, et al. Hepatic cytoprotective effect of ischemic and anesthetic preconditioning before liver resection when using intermittent vascular inflow occlusion: A randomized clinical trial. Surgery (United States). 2015;157(2):249-59.

182. Rogers CA, Pike K, Campbell H, Reeves BC, Angelini GD, Gray A, et al. Coronary artery bypass grafting in high-RISk patients randomised to off- or on-pump surgery: A randomized controlled trial (the CRISP trial). Health Technology Assessment. 2014;18(44):1-157.

183. Rönkä K, Vironen J, Kössi J, Hulmi T, Silvasti S, Hakala T, et al. Randomized multicenter trial comparing glue fixation, self-gripping mesh, and suture fixation of mesh in lichtenstein hernia repair (FinnMesh Study). Annals of Surgery. 2015;262(5):714-20.

184. Rosenfield K, Jaff MR, White CJ, Rocha-Singh K, Mena-Hurtado C, Metzger DC, et al. Trial of a paclitaxel-coated balloon for femoropopliteal artery disease. New England Journal of Medicine. 2015;373(2):145-53.

185. Ruiz-Tovar J, Llavero C, Morales V, Gamallo C. Total Occlusive Ionic Silver-Containing Dressing vs Mupirocin Ointment Application vs Conventional Dressing in Elective Colorectal Surgery: Effect on Incisional Surgical Site Infection. Journal of the American College of Surgeons. 2015;221(2):424-9.

186. Sa Cunha A, Carrere N, Meunier B, Fabre JM, Sauvanet A, Pessaux P, et al. Stump closure reinforcement with absorbable fibrin collagen sealant sponge (TachoSil) does not prevent pancreatic fistula after distal pancreatectomy: The FIABLE multicenter controlled randomized study. American Journal of Surgery. 2015;210(4):739-48.

187. Sadahiro S, Suzuki T, Tanaka A, Okada K, Kamata H, Ozaki T, et al. Comparison between oral antibiotics and probiotics as bowel preparation for elective colon cancer surgery to prevent infection: Prospective randomized trial. Surgery (United States). 2014;155(3):493-503.

188. Salminen P, Paajanen H, Rautio T, Nordström P, Aarnio M, Rantanen T, et al. Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: The

APPAC randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;313(23):2340-8.

189. Sanders DL, Nienhuijs S, Ziprin P, Miserez M, Gingell-Littlejohn M, Smeds S. Randomized clinical trial comparing self-gripping mesh with suture fixation of lightweight polypropylene mesh in open inguinal hernia repair. British Journal of Surgery. 2014;101(11):1373-82.

190. Saver JL, Goyal M, Bonafe A, Diener HC, Levy EI, Pereira VM, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. New England Journal of Medicine. 2015;372(24):2285-95.

191. Saydam M, Ozturk B, Sinan H, Balta AZ, Demir P, Ozer MT, et al. Comparison of modified Limberg flap transposition and lateral advancement flap transposition with Burow's triangle in the treatment of pilonidal sinus disease. American Journal of Surgery. 2015;210(4):772-7.

192. Schauer PR, Bhatt DL, Kirwan JP, Wolski K, Brethauer SA, Navaneethan SD, et al. Bariatric surgery versus intensive medical therapy for diabetes - 3-Year outcomes. New England Journal of Medicine. 2014;370(21):2002-13.

193. Schietroma M, Cecilia EM, Sista F, Carlei F, Pessia B, Amicucci G. High-concentration supplemental perioperative oxygen and surgical site infection following elective colorectal surgery for rectal cancer: A prospective, randomized, double-blind, controlled, single-site trial. American Journal of Surgery. 2014;208(5):719-26.

194. Schietroma M, Piccione F, Cecilia EM, Carlei F, De Santis G, Sista F, et al. How does high-concentration supplemental perioperative oxygen influence surgical outcomes after thyroid surgery? A prospective, randomized, double-blind, controlled, monocentric trial. Journal of the American College of Surgeons. 2015;220(5):921-33.

195. Schoenberger SD, Kim SJ, Shah R, Sheng J, Cherney E. Reduction of interleukin 8 and platelet-derived growth factor levels by topical ketorolac, 0.45%, in patients with diabetic retinopathy. JAMA Ophthalmology. 2014;132(1):32-7.

196. Schultz JK, Yaqub S, Wallon C, Blecic L, Forsmo HM, Folkesson J, et al. Laparoscopic lavage vs primary resection for acute perforated diverticulitis: The SCANDIV randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;314(13):1364-75.

197. Schultz MJ. High versus low positive end-expiratory pressure during general anaesthesia for open abdominal surgery (PROVHILO trial): A multicentre randomised controlled trial. The Lancet. 2014;384(9942):495-503.

198. Shen YM, Chen J, Tian ML, Yang S, Liu SJ, Wang MG. Lichtenstein repair of indirect inguinal hernias with acellular tissue matrix grafts in adolescent patients: A prospective, randomized, controlled trial. Surgery Today. 2014;44(3):429-35.

199. Shyamsundar M, McAuley DF, Shields MO, MacSweeney R, Duffy MJ, Johnston JR, et al. Effect of simvastatin on physiological and biological outcomes in patients undergoing esophagectomy: A randomized placebo-controlled trial. Annals of Surgery. 2014;259(1):26-31.

200. Slaghekke F, Lopriore E, Lewi L, Middeldorp JM, Van Zwet EW, Weingertner AS, et al. Fetoscopic laser coagulation of the vascular equator versus selective coagulation for twin-to-twin transfusion syndrome: An open-label randomised controlled trial. The Lancet. 2014;383(9935):2144-51.

201. Sloothaak DAM, Van Den Berg MW, Dijkgraaf MGW, Fockens P, Tanis PJ, Van Hooft JE, et al. Oncological outcome of malignant colonic obstruction in the Dutch Stent-In 2 trial. British Journal of Surgery. 2014;101(13):1751-7.

202. Smith GE, Barnes R, Chetter IC. Randomized clinical trial of selective versus routine preoperative duplex ultrasound imaging before arteriovenous fistula surgery. British Journal of Surgery. 2014;101(5):469-74.

203. Smith PK, Puskas JD, Ascheim DD, Voisine P, Gelijns AC, Moskowitz AJ, et al. Surgical treatment of moderate ischemic mitral regurgitation. New England Journal of Medicine. 2014;371(23):2178-88.

204. Smith SC, Heal CF, Buttner PG. Prevention of surgical site infection in lower limb skin lesion excisions with single dose oral antibiotic prophylaxis: A prospective randomised placebo-controlled double-blind trial. BMJ Open. 2014;4(7).

205. Sorensen LS, Thorlacius-Ussing O, Schmidt EB, Rasmussen HH, Lundbye-Christensen S, Calder PC, et al. Randomized clinical trial of perioperative omega-3 fatty acid supplements in elective colorectal cancer surgery. British Journal of Surgery. 2014;101(2):33-42.

206. Spruyt O, Westerman D, Milner A, Bressel M, Wein S. A randomised, double-blind, placebo-controlled study to assess the safety and efficacy of methoxyflurane for procedural pain of a bone marrow biopsy. BMJ Supportive and Palliative Care. 2014;4(4):342-8.

207. Sridhar P, Sistla SC, Ali SM, Karthikeyan VS, Badhe AS, Ananthanarayanan PH. Effect of intravenous lignocaine on perioperative stress response and post-surgical ileus in elective open abdominal surgeries: A double-blind randomized controlled trial. ANZ Journal of Surgery. 2015;85(6):425-9.

208. Srinivas A, Kaman L, Raj P, Gautam V, Dahiya D, Singh G, et al. Comparison of the efficacy of chlorhexidine gluconate versus povidone iodine as preoperative skin preparation for the prevention of surgical site infections in clean-contaminated upper abdominal surgeries. Surgery Today. 2015;45(11):1378-84.

209. Steiner ME, Ness PM, Assmann SF, Triulzi DJ, Sloan SR, Delaney M, et al. Effects of red-cell storage duration on patients undergoing cardiac surgery. New England Journal of Medicine. 2015;372(15):1419-29.

210. Stevenson ARL, Solomon MJ, Lumley JW, Hewett P, Clouston AD, Gebski VJ, et al. Effect of laparoscopic-assisted resection vs open resection on pathological outcomes in rectal cancer: The ALaCaRT randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;314(13):1356-63.

211. Sun C, Zhang M, Jiang C. Vertical tunnel-based low-site peritoneal dialysis catheter implantation decreases the incidence of catheter malfunction. American Surgeon. 2015;81(11):1157-62.

212. Svensson JF, Patkova B, Almström M, Naji H, Hall NJ, Eaton S, et al. Nonoperative treatment with antibiotics versus surgery for acute nonperforated appendicitis in children. Annals of Surgery. 2015;261(1):67-71.

213. Takata A, Takiguchi S, Miyazaki Y, Miyata H, Takahashi T, Kurokawa Y, et al. Randomized phase II study of the anti-inflammatory effect of ghrelin during the postoperative period of esophagectomy. Annals of Surgery. 2015;262(2):230-6.

214. Tanaka A, Sadahiro S, Suzuki T, Okada K, Saito G. Randomized controlled trial comparing subcuticular absorbable suture with conventional interrupted suture for wound closure at elective operation of colon cancer. Surgery (United States). 2014;155(3):486-92.

215. Tarantino I, Warschkow R, Beutner U, Kolb W, Luthi A, Luthi C, et al. Efficacy of a single preoperative dexamethasone dose to prevent nausea and vomiting after thyroidectomy (the tPONV Study): A randomized, double-blind, placebo-controlled clinical trial. Annals of Surgery. 2015;262(6):934-40.

216. Tavilla G, Bruggemans EF, Gielen CLI, Brand A, Van Den Hout WB, Klautz RJM, et al. Multicentre randomized clinical trial to investigate the cost-effectiveness of an allogeneic single-donor fibrin sealant after coronary artery bypass grafting (FIBER Study). British Journal of Surgery. 2015;102(11):1338-47.

217. Thanapal MR, Tata MD, Tan AJ, Subramaniam T, Tong JMG, Palayan K, et al. Pre-emptive intraperitoneal local anaesthesia: An effective method in immediate post-operative pain management and metabolic stress response in laparoscopic appendicectomy, a randomized, double-blinded, placebo-controlled study. ANZ Journal of Surgery. 2014;84(1-2):47-51.

218. Tyson AF, Kendig CE, Mabedi C, Cairns BA, Charles AG. The effect of incentive spirometry on postoperative pulmonary function following laparotomy a randomized clinical trial. JAMA Surgery. 2015;150(3):229-36.

219. Valente DS, Steffen N, Carvalho LA, Borille GB, Zanella RK, Padoin AV. Preoperative use of dexamethasone in rhinoplasty a randomized, double-blind, placebo-controlled clinical trial. JAMA Facial Plastic Surgery. 2015;17(3):169-73.

220. Valgimigli M, Gagnor A, Calabró P, Frigoli E, Leonardi S, Zaro T, et al. Radial versus femoral access in patients with acute coronary syndromes undergoing invasive management: A randomised multicentre trial. The Lancet. 2015;385(9986):2465-76.

221. Van Buren G, Bloomston M, Hughes SJ, Winter J, Behrman SW, Zyromski NJ, et al. A randomized prospective multicenter trial of pancreaticoduodenectomy with and without routine intraperitoneal drainage. Annals of Surgery. 2014;259(4):605-12.

222. Van Den Heijkant TC, Costes LMM, Van Der Lee DGC, Aerts B, Osinga-De Jong M, Rutten HRM, et al. Randomized clinical trial of the effect of gum chewing on postoperative ileus and inflammation in colorectal surgery. British Journal of Surgery. 2015;102(3):202-11 and e185.

223. Vather R, Josephson R, Jaung R, Kahokehr A, Sammour T, Bissett I. Gastrografin in prolonged postoperative ileus: A double-blinded randomized controlled trial. Annals of Surgery. 2015;262(1):23-30.

224. Vedovati MC, Becattini C, Rondelli F, Boncompagni M, Camporese G, Balzarotti R, et al. A randomized study on 1-week versus 4-week prophylaxis for venous thromboembolism after laparoscopic surgery for colorectal cancer. Annals of Surgery. 2014;259(4):665-9.

225. Vennix S, Musters GD, Mulder IM, Swank HA, Consten EC, Belgers EH, et al. Laparoscopic peritoneal lavage or sigmoidectomy for perforated diverticulitis with purulent peritonitis: A multicentre, parallel-group, randomised, open-label trial. The Lancet. 2015;386(10000):1269-77.

226. Verheye S, Jolicœur EM, Behan MW, Pettersson T, Sainsbury P, Hill J, et al. Efficacy of a device to narrow the coronary sinus in refractory angina. The New England journal of medicine. 2015;372(6):519-27.

227. Vester-Andersen M, Waldau T, Wetterslev J, Møller MH, Rosenberg J, Jørgensen LN, et al. Randomized multicentre feasibility trial of intermediate care versus standard ward care after emergency abdominal surgery (InCare trial). British Journal of Surgery. 2015;102(6):619-29.

228. Visser N, Beckers HJM, Bauer NJC, Gast STJM, Zijlmans BLM, Berenschot TTJM, et al. Toric vs aspherical control intraocular lenses in patients with cataract and corneal astigmatism: A randomized clinical trial. JAMA Ophthalmology. 2014;132(12):1462-8.

229. Wang C, Zhai H, Liu W, Li J, Yang J, Hu Y, et al. Thyroidectomy: A novel endoscopic oral vestibular approach. Surgery (United States). 2014;155(1):33-8.

230. Ward WG, Cooper JM, Lippert D, Kablawi RO, Neiberg RH, Sherertz RJ. Glove and gown effects on intraoperative bacterial contamination. Annals of Surgery. 2014;259(3):591-7.

231. Watkins AA, Johnson TV, Shrewsberry AB, Nourparvar P, Madni T, Watkins CJ, et al. Ice packs reduce postoperative midline incision pain and narcotic Use: A randomized controlled trial. Journal of the American College of Surgeons. 2014;219(3):511-7.

232. Watson DI, Thompson SK, Devitt PG, Smith L, Woods SD, Aly A, et al. Laparoscopic repair of very large hiatus hernia with sutures versus absorbable mesh versus nonabsorbable mesh a randomized controlled trial. Annals of Surgery. 2015;261(2):282-9.

233. Weissman JD, Most SP. Radiofrequency thermotherapy vs bone-anchored suspension for treatment of lateral nasal wall insufficiency: A randomized clinical trial. JAMA Facial Plastic Surgery. 2015;17(2):84-9.

234. Wentworth JM, Playfair J, Laurie C, Ritchie ME, Brown WA, Burton P, et al. Multidisciplinary diabetes care with and without bariatric surgery in overweight people: A randomised controlled trial. The Lancet Diabetes and Endocrinology. 2014;2(7):545-52.

235. Westin L, Wollert S, Ljungdahl M, Sandblom G, Gunnarsson U, Dahlstrand U. Less pain 1 year after total extra-peritoneal repair compared with lichtenstein using local anesthesia data from a rand omized controlled clinical trial. Annals of Surgery. 2016;263(2):240-3.

236. Weston MK, Stephens JH, Schafer A, Hewett PJ. Warm-up before laparoscopic surgery is not essential. ANZ Journal of Surgery. 2014;84(3):143-7.

237. Whitlock RP, Devereaux PJ, Teoh KH, Lamy A, Vincent J, Pogue J, et al. Methylprednisolone in patients undergoing cardiopulmonary bypass (SIRS): A randomised, double-blind, placebo-controlled trial. The Lancet. 2015;386(10000):1243-53.

238. Wolthuis AM, Fieuws S, Van Den Bosch A, De Buck Van Overstraeten A, D'Hoore A. Randomized clinical trial of laparoscopic colectomy with or without natural-orifice specimen extraction. British Journal of Surgery. 2015;102(6):630-7.

239. Wylde V, Marques E, Artz N, Blom A, Gooberman-Hill R. Effectiveness and cost-effectiveness of a group-based pain self-management intervention for patients undergoing total hip replacement: Feasibility study for a randomized controlled trial. Trials. 2014;15(1).

240. Xu J, Zhang B, Shi S, Qin Y, Ji S, Xu W, et al. Papillary-like main pancreatic duct invaginated pancreaticojejunostomy versus duct-to-mucosa pancreaticojejunostomy after pancreaticoduodenectomy: A prospective randomized trial. Surgery (United States). 2015;158(5):1211-8.

241. Yan M, Ni J, Song D, Ding M, Liu T, Huang J. Radial head replacement or repair for the terrible triad of the elbow: which procedure is better? ANZ Journal of Surgery. 2015;85(9):644-8.

242. Yang C, Chang H, Zhang T, Liang C, Li E. Pre-emptive epidural analgesia improves post-operative pain and immune function in patients undergoing thoracotomy. ANZ Journal of Surgery. 2015;85(6):472-7.

243. Yokoyama Y, Nishigaki E, Abe T, Fukaya M, Asahara T, Nomoto K, et al. Randomized clinical trial of the effect of perioperative synbiotics versus no synbiotics on bacterial translocation after oesophagectomy. British Journal of Surgery. 2014;101(3):189-99.

244. Yu W, Wang W, Rong W, Wang L, Xu Q, Wu F, et al. Adjuvant radiotherapy in centrally located hepatocellular carcinomas after hepatectomy with narrow margin (<1 cm): A prospective randomized study. Journal of the American College of Surgeons. 2014;218(3):381-92.

245. Zaghiyan K, Melmed GY, Berel D, Ovsepyan G, Murrell Z, Fleshner P. A prospective, randomized, noninferiority trial of steroid dosing after major colorectal surgery. Annals of Surgery. 2014;259(1):32-7.

246. Zarbock A, Schmidt C, Van Aken H, Wempe C, Martens S, Zahn PK, et al. Effect of remote ischemic preconditioning on kidney injury among high-risk patients undergoing

cardiac surgery: A randomized clinical trial. JAMA - Journal of the American Medical Association. 2015;313(21):2133-41.

247. Zargar H, Marshall D, Siva G, King Q. Topical diltiazem before transrectal ultrasonography-guided biopsy of the prostate: A randomized controlled trial. ANZ Journal of Surgery. 2015;85(6):430-2.

Appendix 4 – Supplementary material from Chapter 4

Appendix Material 5: Data extraction form

The following details were extracted from eligible funding applications:

SECTION 1: Trial details

1.1. Funding identifier (CATEGORICAL – *EME / HTA*)

1.2. Trial name (FREETEXT)

1.3. Number of randomized controlled trials in application (NUMBERIC)

1.4. Lead institute region (CATEGORICAL – by COUNTY)

1.5. Funding start year (CATEGORICAL – 2012 / 2013 / 2014 / 2015 / 2016 / 2017)

1.6. Documents available for review

1.6.1. Commissioning brief (BINARY – *Yes / No*)

1.6.2. Project description (BINARY – *Yes / No*)

1.6.3. Funder changes (BINARY – *Yes / No*)

1.6.4. Protocol (BINARY – *Yes / No*)

SECTION 2: Design details

2.1. Trial design (CATEGORICAL - *Cluster / Crossover / Parallel / Factorial / Stepped wedge / N-of-1 / Sequential*)

2.2. Number of trial arms (NUMERIC)

2.3. Use of pilot or feasibility in design

2.3.1. Pilot study (BINARY – *Yes / No*)

2.3.2. Feasibility study (BINARY – *Yes / No*)

SECTION 3: Intervention of interest

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

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

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

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

SECTION 4: Recruitment

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

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

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

SECTION 5: Randomisation

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

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

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

5.3. Randomisation unit (BINARY – *Individual / Dyad / Cluster*)

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

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

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

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

SECTION 6: Centre and surgeon credentials

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

6.2. Centre credentials (FREETEXT)

6.3. Surgeon credentials (FREETEXT)

SECTION 7: Outcomes

7.1. Outcomes (FREETEXT)

SECTION 8: Statistical considerations

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

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

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

SECTION 9: Funder led considerations

9.1. Commissioning brief (FREETEXT)

9.2. Funder led changes (FREETEXT)

Appendix 5 – Supplementary material from Chapter 5

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

The following list of UK Clinical Research Collaborative Registered Clinical Trials Units was obtained on 4th January 2019.

1. Barts and the London Pragmatic CTU
2. Barts Clinical Trials Unit
3. Birmingham Clinical Trials Unit
4. Bristol Clinical Trials and Evaluation Unit
5. Bristol Randomised Trials Collaboration
6. CaCTUS (Cancer Clinical Trials Unit Scotland)
7. Cambridge Clinical Trials Unit (CCTU)
8. Cancer Research UK Clinical Trials Unit (CRCTU)
9. Centre for Healthcare Randomised Trials (CHaRT)
10. Centre for Trials Research
11. Comprehensive CTU @ UCL
12. CR UK & UCL Cancer Trials Centre
13. Diabetes Trials Unit (Churchill Hospital, Oxford)
14. Edinburgh Clinical Trials Unit, Edinburgh
15. Glasgow Clinical Trials Unit
16. Imperial Clinical Trials Unit
17. Intensive Care National Audit & Research Centre (ICNARC) CTU
18. Keele Clinical Trials Unit
19. King's Clinical Trials Unit at King's Health Partners
20. Leeds Clinical Trials Research Unit
21. Leicester Clinical Trials Unit
22. Liverpool Trials Collaborative
23. London School of Hygiene & Tropical Medicine
24. Manchester Academic Health Science Centre Clinical Trials Unit (MAHSC-CTU)
25. Medical Research Council Clinical Trials Unit at UCL
26. Newcastle Clinical Trials Unit (NCTU)
27. NHS Blood and Transplant Clinical Trials Unit
28. North Wales Organisation for Randomised Trials in Health (NWORTH)
29. Northern Ireland Clinical Trials Unit

Appendix Box 1 (continued): List of UK Clinical Research Collaborative Registered Clinical Trials Units

30. Norwich Clinical Trials Unit
30. Nottingham Clinical Trials Unit
31. NPEU Clinical Trials Unit
32. Oxford Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU)
33. Oxford Clinical Trials Research Unit (OCTRU)
34. Oxford Primary Care and Vaccines Collaborative Clinical Trials Unit
35. Papworth Trials Unit Collaboration
36. Peninsula Clinical Trials Unit
37. PRIMENT Clinical Trials Unit at UCL
38. Royal Marsden Clinical Trials Unit (RM-CTU)
39. Sheffield Clinical Trials Research Unit
40. Southampton Clinical Trials Unit
41. Swansea Trials Unit
42. Tayside Clinical Trials Unit
43. The Institute of Cancer Research Clinical Trials & Statistics Unit (ICR-CTSU)
44. Warwick Clinical Trials Unit
45. York Trials Unit
46. Brighton and Sussex Clinical Trials Unit
47. Cambridge Epidemiology & Trials Unit
48. Derby Clinical Trials Support Unit (DCTSU)
49. Exeter Clinical Trials Unit
50. Surrey Clinical Trials Unit

Appendix Box 2: Survey

Survey of UK registered Clinical Trials Units

CTU name: _____

Details of CTU representative

Name: _____

Role: _____

Email: _____

All questions within this survey relate to trials with any type of intervention i.e. are not restricted to surgical or pharmacological studies.

Refer to the following definitions when completing this survey

Centre	Designated medical facility or site used to conduct clinical research e.g. hospital or medical clinic.
Treatment provider	The individual e.g. doctor administering the treatment (or intervention) to a consenting patient.
Surgical intervention	Interventions that involve physically changing body tissues and organs through manual operation, such as cutting, abrading, suturing and the use of lasers.[1]

From MRC Guidance for Developing and Evaluating Complex Interventions:

There are fewer trials of surgical than of pharmaceutical interventions, and those there are tend to be of poorer quality. Trials are sometimes dismissed as impossible or inappropriate in surgery, for a variety of reasons. These difficulties include variability in the intervention over time, "learning curve effects", or between surgeons. These affect many other kinds of complex intervention.

A complex intervention is conventionally defined as interventions with several interacting components, they present a number of special problems for evaluators, in addition to the practical and methodological difficulties that any successful evaluation must overcome.

1. Does your unit have experience of running trials with:
   a. A surgical intervention?
      Yes ☐ No ☐
   b. A complex intervention?
      Yes ☐ No ☐
   c. If yes, to a. or b., would you be happy to discuss your answers in more detail following this survey?
      Yes ☐ → *please ensure email address provided above.*
      No ☐

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

1

Appendix Box 2(continued): Survey

The following questions relate to the design of all multicentre trials:

From ICH E9:

Good design should generally aim to achieve the same distribution of subjects to treatments within each centre and good management should maintain this design objective. Trials that avoid excessive variation in the numbers of patients per centre, and trials that avoid a few very small centres, have advantages if it is later found necessary to take into account the heterogeneity of the treatment effect from centre to centre.

2. Does your unit have any multi centre trials that does not stratify randomisation by centre?

   Yes ☐ No ☐

   a. Is this because (tick all that apply):

      i. Expected homogeneity of treatment effect across centres ☐
      ii. No interest in centre effect ☐
      iii. Stratified by treatment provider within centres and treatment providers unique within centre ☐
      iv. Lots of centres with few participants per centre ☐
      v. Not convinced of appropriateness of either fixed or random effect models for centres in this trial ☐
      vi. Other - give details:

      _____
      _____
      _____
      _____
      _____
      _____

2

Appendix Box 2(continued): Survey

3. In the following scenarios has your unit stratified randomisation by:

   a. Scenario A – *A trial with a large [2] sample size, recruiting in several centres each with multiple treatment providers.*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐
      Give details: _____

   b. Scenario B – *A trial with a small [3] sample size, recruiting in several centres each with multiple treatment providers.*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐
      Give details: _____

   c. Scenario C – *A trial recruiting in several centres, where treatment providers work at more than one centre (treatment provider not unique to a centre).*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐
      Give details: _____

   d. Scenario D – *A trial investigating a surgical intervention, recruiting from several centres, each with multiple treatment providers.*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐
      Give details: _____

   e. Scenario E – *A trial investigating substantially different interventions (e.g. surgical vs. non-surgical) recruiting from several centres, each with multiple treatment providers delivering the interventions.*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐
      Give details: _____

[2] With centres recruiting at least ten patients per site.
[3] With centres recruiting 2-3 patients per site.

3

Appendix Box 2(continued): Survey

From ICH E9:

In cases of doubt the aim should be to define centres so as to achieve homogeneity in the important factors affecting the measurements of the primary variables and the influence of the treatments. Any rules for combining centres in the analysis should be justified and specified prospectively in the protocol where possible.

4. Has your unit combined **(a) centres** or **(b) treatment providers** for the purpose of analyses?

	(a)	(b)
Yes, pre-specified grouping rules at design stage	☐	☐
Yes, ad hoc approach e.g. determined after design due to small numbers per group	☐	☐
No, have not combined groups	☐	☐
Not applicable, no experience of trials of this type		☐

Other - give details: _____

_____

_____

_____

One approach is to increase homogeneity of centre or treatment provider effects by standardising training. The following questions are related to centre or treatment provider requirements and training provisions.

From ICH E9:

It is important to design the common protocol and to conduct the trial with this background in mind. Procedures should be standardised as completely as possible.

5. Has your unit defined a minimum level of experience for participating treatment providers in terms of:
   a) Treating the condition within the patient population?
      Yes, when applicable ☐ No, with justification ☐ No ☐
   b) Delivering the trial intervention?
      Yes, when applicable ☐ No, with justification ☐ No ☐
   c) Setting a minimum professional level of treatment providers e.g. minimum speciality training level
      Yes, when applicable ☐ No, with justification ☐ No ☐
   d) Other – give details: _____

      _____

      _____

      _____

6. Has your unit conducted trials with an expertise based design, in which participating treatment providers provide only the intervention in which they have expertise?

   Yes, when applicable ☐ No, with justification ☐ No ☐

4

Appendix Box 2 (continued): Survey

The following questions are related to analysis of multicentre trials

From ICH E9:

Factors on which randomisation has been stratified should be accounted for later in the analysis.

When analysing a trial, failure to have balance within groups, combined with doubts about the homogeneity of the results may, in severe cases, reduce the value of a multicentre trial to such a degree that it cannot be regarded as giving convincing evidence for the sponsor's claims.

7. Does your unit include centre in the statistical model when comparing treatment?
   - Yes, but only if it was used to stratify randomisation ☐
   - Yes, always ☐
   - Yes, sometimes ☐
   - No, never ☐

   a. If yes, is this effect treated as fixed or random?
   Fixed ☐ Random ☐ Depends ☐

   b. Give details regarding your answer

   _____
   _____
   _____
   _____

8. Does your unit include treatment provider in the statistical model when comparing treatment?
   - Yes, but only if it was used to stratify randomisation ☐
   - Yes, always ☐
   - Yes, sometimes ☐
   - No, never ☐

   a. If yes, is this treated as fixed or random effects?
   Fixed ☐ Random ☐ Depends ☐
   Give details regarding your answer

   _____
   _____
   _____
   _____

   b. Have you ever treated the effect of the treatment provider as time varying? For example, allowed for expertise to change over time?
   Yes ☐ No ☐
   Give details regarding your answer

   _____
   _____
   _____
   _____

5

Appendix Box 2 (continued): Survey

From ICH E9:

The statistical model to be adopted for the estimation and testing of treatment effects should be described in the protocol. The main treatment effect may be investigated first using a model which allows for centre differences, but does not include a term for treatment-by-centre interaction. If the treatment effect is homogenous across centres, the routing inclusion of interaction terms in the model reduces the efficiency of the test for the main effects. In the presence of true heterogeneity of treatment effects, the interpretation of the main treatment effect is controversial.

9. Regardless of your randomisation stratification approach, does your unit check for a treatment by centre and/or treatment provider interaction effect in the following scenarios?
   a. Scenario A – *A trial with a large [2] sample size, recruiting in several centres each with multiple treatment providers.*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐

      Give details: _____

   b. Scenario B – *A trial with a small [3] sample size, recruiting in several centres each with multiple treatment providers.*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐

      Give details: _____

   c. Scenario C – *A trial recruiting in several centres, where treatment providers work at more than one centre (treatment provider not unique to a centre).*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐

      Give details: _____

   d. Scenario D – *A trial investigating a surgical intervention, recruiting from several centres, each with multiple treatment providers.*

      Centre ☐ Treatment provider ☐
      Both ☐ Neither ☐
      N/A – no experience of trials of this type ☐

      Give details: _____

[2] With centres recruiting at least ten patients per site.
[3] With centres recruiting 2-3 patients per site.

6

Appendix Box 2 (continued): Survey

(9. cont.) Regardless of your randomisation stratification approach, does your unit check for a treatment by centre and/or treatment provider interaction effect in the following scenarios?

e. Scenario E – *A trial investigating substantially different interventions, e.g. surgical vs. non-surgical, recruiting from several centres, each with multiple treatment providers delivering the interventions.*

Centre ☐ Treatment provider ☐
Both ☐ Neither ☐
N/A – no experience of trials of this type ☐

Give details: _____

From ICH E9:

If positive treatment effects are found in a trial with appreciable numbers of subjects per centre there should generally be an exploration of the heterogeneity of treatment effects across centres, as this may affect the generalisability of the conclusions. Marked heterogeneity may be identified by graphical display of the results of individual centres or by analytical methods, such as a significance test of the treatment-by-centre interaction. When using such a statistical significance test, it is important to recognise that this generally has low power in a trial designed to detect the main effect of treatment.

10. If a positive treatment effect is found, does your unit, explore heterogeneity of treatment effects by:

a. Centre

Yes ☐ No ☐ Depends ☐

If yes, how? Select all that apply.

i. Graphical display or results of individual centres
Yes ☐ No ☐

ii. Analytical methods e.g. significance test of treatment-by-centre interaction
Yes ☐ No ☐

Give details:

_____

b. Treatment provider – unique and not unique between centres

Yes ☐ No ☐

If yes, how? Select all that apply.

i. Graphical display or results of individual centres
Yes ☐ No ☐

ii. Analytical methods e.g. significance test of treatment-by-treatment provider interaction
Yes ☐ No ☐

Give details:

_____

7

Appendix Box 2 (continued): Survey

Additional notes page

Please use the following to provide more details responses as required.

8

Appendix Table 2: 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	

ID	Scenario	Stratification approaches	Free text
	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	For D, centre: Hospital; Treatment provider: operating surgeon, closing the wound.
		Treatment provider only	
		Both	
	E	No experience	
ID13	A	Centre only	We rarely collect data on treatment provider.
	B	Centre only	
	C	No experience	
ID14	A	Centre only	For a, dependent on trial. General comments - considered and decided on a trial by trial basis. Don't feel you can always standardise.
		Both	
	B	Centre only	
	C	Treatment provider only	
	D	Centre only	General comments - considered and decided on a trial by trial basis. Don't feel you can always standardise.
		Treatment provider only	
	E	Centre only	

ID	Scenario	Stratification approaches	Free text
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 as their argument for including. For D and E, not aware of any locally, but if we did then think definitely by treatment provider.
	B	Neither centre nor treatment provider	
	C	No experience	
	D	No experience	
	E	No experience	
ID29	A	Both	For C, stratification is likely to be chosen by the expected homogeneity, so may be
	B	Centre only	
	C	Centre only	

ID	Scenario	Stratification approaches	Free text
		Treatment provider only	centre or treatment provider. This will be intervention specific.
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, [*Name of* trial] Surgeons within seven centres. Surgery is conducted by a team which includes a variable subset of
	B	Centre only	
	C	No experience	

ID	Scenario	Stratification approaches	Free text
			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.
	D	Centre only	For D, [*Name of* trial] 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, [*Name of* trial] 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	

Appendix 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.

Appendix 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 out 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 multilevel 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.

Appendix Table 3: 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	

ID	Scenario	Stratification approaches	Free text
	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.

ID	Scenario	Stratification approaches	Free text
	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.

ID	Scenario	Stratification approaches	Free text
	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 the 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	

ID	Scenario	Stratification approaches	Free text
	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

ID	Scenario	Stratification approaches	Free text
	E	Neither centre nor treatment provider	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.

Appendix Table 4: 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

ID	Graphical display	Analytical methods	Further details on exploring heterogeneity by centre
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.

ID	Graphical display	Analytical methods	Further details on exploring heterogeneity by centre
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

Appendix Table 5: 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.

Appendix 6 – Supplementary material from Chapter 7

Appendix Material 6: SAS macro to estimate VPC

```
%MACRO ICC(U=Covariance estimate for random effect, INT=Intercept estimate,
TRT=Treatment effect estimate);
DATA ICCSIM;
CALL streaminit(25345278);
DO SET=1 TO 5000;
U=RAND('normal',0,SQRT(&U));
LOGITP=&INT+&TRT+U;
P=EXP(LOGITP)/((1+EXP(LOGITP)));
V=P*(1-P);
OUTPUT;
END;
RUN;

proc means data=ICCSIM;
var V;
output out=S mean=;
run;

proc means data=ICCSIM;
var P;
output out=VAR_P VAR=;
run;

DATA ICCEST;
```

```
MERGE S VAR_P;
BY _TYPE_ _FREQ_;
RUN;

DATA ICCEST;
SET ICCEST;
KEEP VPC;
VPC=P/(P+V);
RUN;
%mend ICC;
```

Appendix Table 6: Summary of failure for any cause by centre in BASICS

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
Centre 1	Shunt inserted	188	62	64	62
	Failure (%)	18 (9.6%)	4 (6.5%)	7 (10.9%)	7 (11.3%)
	[97.5% CI]	[4.8, 14.4]	[0.0, 13.4]	[2.2, 19.7]	[2.3, 20.3]
Centre 2	Shunt inserted	175	58	59	58
	Failure (%)	35 (20.0%)	9 (15.5%)	12 (20.3%)	14 (24.1%)
	[97.5% CI]	[13.2, 26.8]	[4.9, 26.2]	[8.6, 32.1]	[11.6, 36.7]
Centre 3	Shunt inserted	154	53	52	49
	Failure (%)	40 (26.0%)	13 (24.5%)	15 (28.8%)	12 (24.5%)
	[97.5% CI]	[18.1, 33.9]	[11.3, 37.8]	[14.8, 42.9]	[10.7, 38.3]
Centre 4	Shunt inserted	140	47	45	48
	Failure (%)	23 (16.4%)	4 (8.5%)	11 (24.4%)	8 (16.7%)
	[97.5% CI]	[9.4, 23.4]	[0.0, 17.6]	[10.1, 38.8]	[4.6, 28.7]
Centre 5	Shunt inserted	128	43	44	41
	Failure (%)	30 (23.4%)	9 (20.9%)	10 (22.7%)	11 (26.8%)
	[97.5% CI]	[15.1, 31.8]	[7.0, 34.8]	[8.6, 36.9]	[11.3, 42.3]
Centre 6	Shunt inserted	118	39	40	39
	Failure (%)	56 (47.5%)	19 (48.7%)	18 (45.0%)	19 (48.7%)
	[97.5% CI]	[37.2, 57.8]	[30.8, 66.6]	[27.4, 62.6%]	[30.8, 66.6]
Centre 7	Shunt inserted	91	30	31	30
	Failure (%)	21 (23.1%)	9 (30.0%)	5 (16.1%)	7 (23.3%)
	[97.5% CI]	[13.2, 33.0]	[11.3, 48.7]	[1.3, 30.9]	[6.0, 40.6]

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
Centre 8	Shunt inserted	84	28	27	29
	Failure (%)	12 (14.3%)	5 (17.9%)	3 (11.1%)	4 (13.8%)
	[97.5% CI]	[5.7, 22.8]	[1.6, 34.0]	[0.0, 24.7]	[0.0, 28.1]
Centre 9	Shunt inserted	81	27	26	28
	Failure (%)	25 (30.1%)	13 (48.1%)	5 (19.2%)	7 (25.0%)
	[97.5% CI]	[19.4, 42.4]	[26.6, 69.7]	[1.9, 36.5]	[6.7, 43.3]
Centre 10	Shunt inserted	73	25	24	24
	Failure (%)	8 (11.0%)	2 (8.0%)	3 (12.5%)	3 (12.5%)
	[97.5% CI]	[2.8, 19.1]	[0.0, 20.2]	[0.0, 27.6]	[0.0, 27.6]
Centre 11	Shunt inserted	71	24	23	24
	Failure (%)	23 (32.4%)	8 (33.3%)	6 (26.1%)	9 (37.5%)
	[97.5% CI]	[20.0, 44.8]	[11.8, 54.9]	[5.6, 46.6]	[15.4, 59.6]
Centre 12	Shunt inserted	68	23	23	22
	Failure (%)	34 (50.0%)	13 (56.5%)	13 (56.5%)	8 (36.4%)
	[97.5% CI]	[36.4, 63.6]	[33.4, 79.7]	[33.4, 79.7]	[13.4, 59.3]
Centre 13	Shunt inserted	47	16	16	15
	Failure (%)	20 (42.5%)	5 (31.3%)	8 (50.0%)	7 (46.7%)
	[97.5% CI]	[26.4, 58.7]	[5.3, 57.2]	[22.0, 78.0]	[17.8, 75.5]
Centre 14	Shunt inserted	40	13	15	12
	Failure (%)	19 (47.5%)	7 (53.8%)	7 (46.7%)	5 (41.7%)
	[97.5% CI]	[29.8, 65.2]	[22.9, 84.8]	[17.8, 75.5]	[9.8, 73.5]
Centre 15	Shunt inserted	36	10	13	13

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
	Failure (%)	6 (16.7%)	1 (10.0%)	3 (23.1%)	2 (15.4%)
	[97.5% CI]	[2.8, 30.6]	[0.0, 31.3]	[0.0, 49.3]	[0.0, 37.8]
Centre 16	Shunt inserted	30	10	10	10
	Failure (%)	8 (26.7%)	1 (10.0%)	3 (30.0%)	4 (40.0%)
	[97.5% CI]	[8.6, 44.8]	[0.0, 31.3]	[0.0, 62.5]	[5.3, 74.7]
Centre 17	Shunt inserted	22	7	8	7
	Failure (%)	9 (40.1%)	5 (71.4%)	2 (25.0%)	2 (28.6%)
	[97.5% CI]	[17.4, 64.4]	[33.2, 100.0]	[0.0, 59.3]	[0.0, 66.8]
Centre 18	Shunt inserted	21	8	7	6
	Failure (%)	1 (4.8%)	0 (0.0%)	0 (0.0%)	1 (16.7%)
	[97.5% CI]	[0.0, 15.2]	[., .]	[., .]	[0.0, 50.7]
Centre 19	Shunt inserted	14	5	5	4
	Failure (%)	2 (14.3%)	0 (0.0%)	0 (0.0%)	2 (50.0%)
	[97.5% CI]	[0.0, 35.2]	[., .]	[., .]	[0.0, 100.0]
Centre 20	Shunt inserted	8	3	2	3
	Failure (%)	6 (75.0%)	2 (66.7%)	1 (50.0%)	3 (100.0%)
	[97.5% CI]	[40.7, 100.0]	[5.7, 100.0]	[0.0, 100.0]	[., .]
Centre 21	Shunt inserted	5	2	1	2
	Failure (%)	2 (40.0%)	1 (50.0%)	0 (0.0%)	1 (50.0%)
	[97.5% CI]	[0.0, 89.1]	[0.0, 100.0]	[., .]	[0.0, 100.0]

Appendix Table 7: Summary of failure for any cause by surgeon in BASICS

Table includes surgeons who operated on at least ten patients only. Table sorted by total number of operations per surgeon.

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
Surgeon 1	Shunt inserted	129	45	40	44
	Failure (%)	8 (6.2%)	3 (6. 7%)	2 (5%)	3 (6.8%)
	[97.5% CI]	[1.5, 11.0]	[0, 15.0]	[0,12.7]	[0, 15.3]
Surgeon 3	Shunt inserted	49	19	18	12
	Failure (%)	8 (16.3%)	4 (21.1%)	4 (22.2%)	0 (0%)
	[97.5% CI]	[4.5, 28.2]	[0.1, 42.0]	[0.3, 44.2]	[0, 0]
Surgeon 6	Shunt inserted	39	13	13	13
	Failure (%)	3 (7.7%)	0 (0%)	1 (7.7%)	2(15.4%)
	[97.5% CI]	[0.0, 17.3]	[., .]	[0, 24.2]	[0, 37.8]
Surgeon 7	Shunt inserted	38	14	12	12
	Failure (%)	15 (39.5%)	4 (28.6%)	5 (41.7%)	6 (50.0%)
	[97.5% CI]	[21.7, 57.2]	[1.5, 55.6]	[9.8, 73.6]	[17.7, 82.3]
Surgeon 4	Shunt inserted	38	18	12	8
	Failure (%)	1 (2.6%)	0 (0%)	1 (8.3%)	0 (0%)
	[97.5% CI]	[0.0, 8.4]	[., .]	[0, 26.2]	[., .]
Surgeon 10	Shunt inserted	33	10	14	9
	Failure (%)	14 (42.4%)	4 (40.0%)	7 (50.0%)	3 (33.3%)
	[97.5% CI]	[23.2, 61.7]	[5.3, 74.7]	[20.1, 79.9]	[0, 68.5]
Surgeon 5	Shunt inserted	32	9	13	10
	Failure (%)	8 (25.0%)	2 (22.2%)	3 (23.1%)	3 (30.0%)

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
	[97.5% CI]	[7.8, 4]	[0, 53.3]	[0, 49.3]	[0, 62.5]
Surgeon 13	Shunt inserted	30	8	12	10
	Failure (%)	9 (30.0%)	4 (50.0%)	2 (16.7%)	3 (30.0%)
	[97.5% CI]	[11.3, 48.7]	[10.4, 89.6]	[0.0, 40.8]	[0, 62.5]
Surgeon 9	Shunt inserted	30	13	7	10
	Failure (%)	5 (16.7%)	2 (15.4%)	1 (14.3%)	2 (20.0%)
	[97.5% CI]	[1.4, 31.9]	[0.0, 37.8]	[0, 43.9]	[0, 48.3]
Surgeon 2	Shunt inserted	25	7	5	13
	Failure (%)	11 (44.0%)	3 (42.9%)	2 (40.0%)	6 (46.2%)
	[97.5% CI]	[21.8, 66.2]	[1.0, 84.8]	[0.0, 89.1]	[15.2, 77.1]
Surgeon 12	Shunt inserted	21	5	8	8
	Failure (%)	4 (19.0%)	0 (0.0%)	1 (12.5%)	3 (37.5%)
	[97.5% CI]	[0.0, 38.2]	[., .]	[0.0, 38.7]	[0, 75.8]
Surgeon 19	Shunt inserted	20	6	8	6
	Failure (%)	3 (15.0%)	0 (0%)	1 (12.5%)	2 (33.3%)
	[97.5% CI]	[0.0, 32.9]	[., .]	[0, 38.7]	[0, 76.4]
Surgeon 8	Shunt inserted	19	7	5	7
	Failure (%)	2 (10.5%)	1 (14.3%)	0 (0.0%)	1 (14.3%)
	[97.5% CI]	[0.0, 26.3]	[0.0, 43.9]	[., .]	[0.0, 43.9]
Surgeon 30	Shunt inserted	19	9	5	5
	Failure (%)	1 (5.3%)	0 (0.0%)	1 (20.0%)	0 (0%)
	[97.5% CI]	[0.0, 16.7]	[., .]	[0, 60.1]	[., .]

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
Surgeon 24	Shunt inserted	18	6	7	5
	Failure (%)	3 (16.7%)	1 (16. 7%)	1 (14.3%)	1 (20.0%)
	[97.5% CI]	[0.0, 36.3]	[0, 50.7]	[0, 43.9]	[0, 60.1]
Surgeon 26	Shunt inserted	17	9	3	5
	Failure (%)	10 (58.8%)	5 (55.6%)	2 (66.7%)	3 (60.0%)
	[97.5% CI]	[32.1, 85.6]	[18.5, 92.7]	[5.7, 100.0]	[10.9, 100.0]
Surgeon 18	Shunt inserted	17	6	7	4
	Failure (%)	4 (23.5%)	1 (16.7%)	1 (14.3%)	2 (50.0%)
	[97.5% CI]	[0.5, 46.6]	[0, 50.8]	[0, 43.9]	[0.0, 100.0]
Surgeon 38	Shunt inserted	16	2	4	10
	Failure (%)	11 (68.8%)	2 (100.0%)	3 (75.0%)	6 (60.0%)
	[97.5% CI]	[42.8, 94.7]	[100.0, 100.0]	[26.5, 100.0]	[25.3, 94.7]
Surgeon 46	Shunt inserted	16	4	8	4
	Failure (%)	5 (31.3%)	1 (25.0%)	3 (37.5%)	1 (25.0%)
	[97.5% CI]	[5.3, 57.2]	[0, 73.5]	[0, 75.8]	[0, 73.5]
Surgeon 34	Shunt inserted	16	7	4	5
	Failure (%)	2 (12.5%)	0 (0.0%)	0 (0.0%)	2 (40.0%)
	[97.5% CI]	[0.0, 31.0]	[., .]	[., .]	[0, 89.1]
Surgeon 21	Shunt inserted	15	3	6	6
	Failure (%)	6 (40.0%)	0 (0%)	2 (33.3%)	4 (66.7%)
	[97.5% CI]	[11.7, 68.3]	[., .]	[0, 76.4]	[23.6, 100.0]

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
Surgeon 37	Shunt inserted	15	3	6	6
	Failure (%)	4 (26.7%)	1 (33.3%)	3 (50.0%)	0 (0.0%)
	[97.5% CI]	[1.1, 52.2]	[0, 94.3]	[4.3, 95.7]	[., .]
Surgeon 15	Shunt inserted	15	6	5	4
	Failure (%)	2 (13.3%)	1 (16. 7%)	0 (0.0%)	1 (25.0%)
	[97.5% CI]	[0.0, 33.0]	[0, 50.7]	[., .]	[0, 73.5]
Surgeon 45	Shunt inserted	15	5	4	6
	Failure (%)	2 (13.3%)	1 (20%)	0 (0.0%)	1 (16.7%)
	[97.5% CI]	[0.0, 33.0]	[0.1, 60.1]	[., .]	[0, 50.7]
Surgeon 23	Shunt inserted	14	4	6	4
	Failure (%)	5 (35.7%)	2 (50.0%)	2 (33.3%)	1 (25.0%)
	[97.5% CI]	[7.1, 64.4]	[0.0, 100.0]	[0, 76.4]	[0, 73.5]
Surgeon 11	Shunt inserted	14	4	2	8
	Failure (%)	2 (14.3%)	1 (25.0%)	0 (0%)	1 (12.5%)
	[97.5% CI]	[0.0, 35.2]	[0, 73.5]	[., .]	[0.0, 38.7]
Surgeon 25	Shunt inserted	13	2	8	3
	Failure (%)	7 (53.8%)	1 (50.0%)	3 (37.5%)	3 (100.0%)
	[97.5% CI]	[22.9, 84.8]	[0.0, 100.0]	[0.0, 75.8]	[100, 100]
Surgeon 63	Shunt inserted	13	7	3	3
	Failure (%)	3 (23.1%)	2 (28.6%)	1 (33.3%)	0 (0.0%)
	[97.5% CI]	[0.0, 49.3]	[0, 66.8]	[0, 94.3]	[., .]
Surgeon 20	Shunt inserted	13	6	3	4

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
	Failure (%)	0 (0.0%)	0 (0%)	0 (0%)	0 (0%)
	[97.5% CI]	[., .]	[., .]	[., .]	[., .]
Surgeon 29	Shunt inserted	12	4	3	5
	Failure (%)	4 (33.3%)	1 (25.0%)	1(33.3%)	2 (40.0%)
	[97.5% CI]	[2.9, 63.8]	[0, 73.5]	[0, 94.3]	[0, 89.1]
Surgeon 17	Shunt inserted	11	1	4	6
	Failure (%)	6 (54.5%)	0 (0%)	2 (50.0%)	4 (66.7%)
	[97.5% CI]	[20.9, 88.2]	[., .]	[0.0, 100.0]	[23.6, 100.0]
Surgeon 54	Shunt inserted	11	3	5	3
	Failure (%)	6 (54.5%)	2 (66.7%)	3 (60.0%)	1(33.3%)
	[97.5% CI]	[20.9, 88.2]	[5.7, 100.0]	[10.9, 100.0]	[0, 94.3]
Surgeon 44	Shunt inserted	11	0	4	7
	Failure (%)	5 (45.5%)	. (%)	1 (25.0%)	4 (57.1%)
	[97.5% CI]	[11.8, 79.1]	[,]	[0, 73.5]	[15.2, 99.0]
Surgeon 16	Shunt inserted	11	3	3	5
	Failure (%)	3 (27.3%)	0 (0.0%)	1 (33.3%)	2 (40.0%)
	[97.5% CI]	[0.0, 57.4]	[., .]	[0.0, 94.3]	[0, 89.1]
Surgeon 43	Shunt inserted	11	3	3	5
	Failure (%)	3 (27.3%)	0 (0%)	1(33.3%)	2 (40.0%)
	[97.5% CI]	[0.0, 57.4]	[., .]	[0, 94.3]	[0, 89.1]
Surgeon 56	Shunt inserted	11	5	5	1
	Failure (%)	3 (27.3%)	1 (20.0%)	1 (20.0%)	1 (100.0%)

		Failure for any cause			
		Overall	Standard	Antibiotic	Silver
Overall	Shunt inserted	1594	533	535	526
	Failure (%)	398 (25.0%)	130 (24.4%)	132 (24.7%)	136 (25.9%)
	[97.5% CI]	[22.5, 27.4]	[20.2, 28.6]	[20.5, 28.8]	[21.6, 30.1]
	[97.5% CI]	[0.0, 57.4]	[0, 60.1]	[0, 60.1]	[100.0, 100.0]
Surgeon 33	Shunt inserted	11	2	4	5
	Failure (%)	1 (9.1%)	0 (0%)	1 (25.0%)	0 (0.0%)
	[97.5% CI]	[0.0, 28.5]	[0.0, 0.0]	[0, 73.5]	[., .]
Surgeon 55	Shunt inserted	11	3	4	4
	Failure (%)	1 (9.1%)	1 (33.3%)	0 (0.0%)	0 (0.0%)
	[97.5% CI]	[0.0, 28.5]	[0, 94.3]	[., .]	[., .]
Surgeon 72	Shunt inserted	10	3	4	3
	Failure (%)	3 (30.0%)	1 (33.3%)	0 (0.0%)	2 (66.7%)
	[97.5% CI]	[0.0, 62.5]	[0, 94.3]	[., .]	[5.7, 100.0]
Surgeon 31	Shunt inserted	10	4	3	3
	Failure (%)	2 (20.0%)	1 (25.0%)	1 (33.3%)	0 (0.0%)
	[97.5% CI]	[0.0, 48.3]	[0, 73.5]	[0, 94.3]	[., .]
Surgeon 32	Shunt inserted	10	4	2	4
	Failure (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[97.5% CI]	[., .]	[., .]	[., .]	[., .]

Appendix Table 8: Summary of fistula by centre in TOPS

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
Centre 1	Surgeries	165	83	82
	Fistula (%)	16 (9.7%)	8 (9.6%)	8 (9.8%)
	[95% CI]	[5.2, 14.2]	[3.3, 16.0]	[3.3, 16.2]
Centre 2	Surgeries	45	23	22
	Fistula (%)	8 (17.8%)	4 (17.4%)	4 (18.2%)
	[95% CI]	[6.6, 28.9]	[1.9, 32.9]	[2.1, 34.3]
Centre 3	Surgeries	41	22	19
	Fistula (%)	8 (19.5%)	5 (22.7%)	3 (15.8%)
	[95% CI]	[7.4, 31.6]	[5.2, 40.2]	[0.0, 32.2]
Centre 4	Surgeries	37	19	18
	Fistula (%)	7 (18.9%)	4 (21.1%)	3 (16.7%)
	[95% CI]	[6.3, 31.5]	[2.7, 39.4]	[0.0, 33.9]
Centre 5	Surgeries	30	15	15
	Fistula (%)	3 (10.0%)	1 (6.7%)	2 (13.3%)
	[95% CI]	[0.0, 20.7]	[0.0, 19.3]	[0.0, 30.5]
Centre 6	Surgeries	27	13	14
	Fistula (%)	2 (7.4%)	1 (7.7%)	1 (7.1%)
	[95% CI]	[0.0, 17.3]	[0.0, 22.2]	[0.0, 20.6]
Centre 7	Surgeries	26	14	12
	Fistula (%)	3 (11.5%)	2 (14.3%)	1 (8.3%)
	[95% CI]	[0.0, 23.8]	[0.0, 32.6]	[0.0, 24.0]

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
Centre 8	Surgeries	24	13	11
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Centre 9	Surgeries	24	11	13
	Fistula (%)	4 (16.7%)	1 (9.1%)	3 (23.1%)
	[95% CI]	[1.8, 31.6]	[0.0, 26.1]	[0.2, 46.0]
Centre 10	Surgeries	22	12	10
	Fistula (%)	2 (9.1%)	2 (16.7%)	0 (0.0%)
	[95% CI]	[0.0, 21.1]	[0.0, 37.8]	[., .]
Centre 11	Surgeries	19	10	9
	Fistula (%)	6 (31.6%)	2 (20.0%)	4 (44.4%)
	[95% CI]	[10.7, 52.5]	[0.0, 44.8]	[12.0, 76.9]
Centre 12	Surgeries	15	6	9
	Fistula (%)	3 (20.0%)	2 (33.3%)	1 (11.1%)
	[95% CI]	[0.0, 40.2]	[0.0, 71.1]	[0.0, 31.6]
Centre 13	Surgeries	12	6	6
	Fistula (%)	1 (8.3%)	1 (16.7%)	0 (0.0%)
	[95% CI]		[0.0, 46.5]	[., .]
Centre 14	Surgeries	12	5	7
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Centre 15	Surgeries	11	6	5

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
	Fistula (%)	3 (27.3%)	3 (50.0%)	0 (0.0%)
	[95% CI]	[1.0, 53.6]	[10.0, 90.0]	[., .]
Centre 16	Surgeries	9	5	4
	Fistula (%)	1 (11.1%)	1 (20.0%)	0 (0.0%)
	[95% CI]	[0.0, 31.6]	[0.0, 55.1]	[., .]
Centre 17	Surgeries	8	5	3
	Fistula (%)	1 (12.5%)	1 (20.0%)	0 (0.0%)
	[95% CI]	[0.0, 35.4]	[0.0, 55.1]	[., .]
Centre 18	Surgeries	7	3	4
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Centre 19	Surgeries	6	3	3
	Fistula (%)	3 (50.0%)	1 (33.3%)	2 (66.7%)
	[95% CI]	[10.0, 90.0]	[0.0, 86.7]	[13.3, 100.0]
Centre 20	Surgeries	5	2	3
	Fistula (%)	2 (40.0%)	1 (50.0%)	1 (33.3%)
	[95% CI]	[0.0, 82.9]	[0.0, 100.0]	[0.0, 86.7]
Centre 21	Surgeries	4	1	3
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Centre 22	Surgeries	3	2	1
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
	[95% CI]	[., .]	[., .]	[., .]

Appendix Table 9: Summary of fistula by surgeon in TOPS

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
Surgeon 1	Surgeries	69	30	39
	Fistula (%)	9 (13.0%)	4 (13.3%)	5 (12.8%)
	[95% CI]	[5.1, 21.0]	[1.2, 25.5]	[2.3, 23.3]
Surgeon 2	Surgeries	11	5	6
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 3	Surgeries	37	19	18
	Fistula (%)	7 (18.9%)	4 (21.1%)	3 (16.7%)
	[95% CI]	[6.3, 31.5]	[2.7, 39.4]	[0.0, 33.9]
Surgeon 4	Surgeries	41	22	19
	Fistula (%)	8 (19.5%)	5 (22.7%)	3 (15.8%)
	[95% CI]	[7.4. 31.6]	[5.2, 40.2]	[0.0, 32.2]
Surgeon 5	Surgeries	22	12	10
	Fistula (%)	2 (9.1%)	2 (16.7%)	0 (0.0%)
	[95% CI]	[0.0, 21.1]	[0.0, 37.8]	[., .]
Surgeon 6	Surgeries	4	1	3
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 7	Surgeries	19	10	9
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
Surgeon 8	Surgeries	11	6	5
	Fistula (%)	3 (27.3%)	3 (50.0%)	0 (0.0%)
	[95% CI]	[1.0, 53.6]	[10.0, 90.0]	[., .]
Surgeon 9	Surgeries	21	13	8
	Fistula (%)	2 (9.5%)	2 (15.4%)	0 (0.0%)
	[95% CI]	[0.0, 22.1]	[0.0, 35.0]	[., .]
Surgeon 10	Surgeries	85	44	41
	Fistula (%)	7 (8.2%)	4 (9.1%)	3 (7.3%)
	[95% CI]	[2.4, 14.1]	[0.6, 17.6]	[0.0, 15.3]
Surgeon 11	Surgeries	6	2	4
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 12	Surgeries	22	11	11
	Fistula (%)	4 (18.2%)	1 (9.1%)	3 (27.3%)
	[95% CI]	[2.1, 34.3]	[0.0, 26.1]	[1.0, 53.6]
Surgeon 13	Surgeries	20	12	8
	Fistula (%)	2 (10.0%)	1 (8.3%)	1 (12.5%)
	[95% CI]	[0.0, 23.1]	[0.0, 24.0]	[0.0, 35.4]
Surgeon 14	Surgeries	6	3	3
	Fistula (%)	3 (50.0%)	1 (33.3%)	2 (66.7%)
	[95% CI]	[10.0, 90.0]	[0.0, 86.7]	[13.3, 100.0]
Surgeon 15	Surgeries	5	1	4

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
	Fistula (%)	1 (20.0%)	0 (0.0%)	1 (25.0%)
	[95% CI]	[0.0, 55.1]	[., .]	[0.0, 67.4]
Surgeon 16	Surgeries	14	6	8
	Fistula (%)	3 (21.4%)	2 (33.3%)	1 (12.5%)
	[95% CI]	[0.0, 42.9]	[0.0, 71.1]	[0.0, 35.4]
Surgeon 17	Surgeries	25	12	13
	Fistula (%)	2 (8.0%)	1 (8.3%)	1 (7.7%)
	[95% CI]	[0.0, 18.6]	[0.0, 24.0]	[0.0, 22.2]
Surgeon 18	Surgeries	20	10	10
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 19	Surgeries	22	12	10
	Fistula (%)	6 (27.3%)	2 (16.7%)	4 (40.0%)
	[95% CI]	[8.7, 45.9]	[0.0, 37.8]	[9.6, 70.4]
Surgeon 20	Surgeries	8	5	3
	Fistula (%)	3 (37.5%)	2 (40.0%)	1 (33.3%)
	[95% CI]	[4.0, 71.0]	[0.0, 82.9]	[0.0, 86.7]
Surgeon 21	Surgeries	17	9	8
	Fistula (%)	5 (29.4%)	2 (22.2%)	3 (37.5%)
	[95% CI]	[7.8, 51.1]	[0.0, 49.4]	[4.0, 71.0]
Surgeon 22	Surgeries	7	3	4
	Fistula (%)	2 (28.6%)	1 (33.3%)	1 (25.0%)

		Fistula		
		Overall	Six months	Twelve months
Overall	Surgeries	552	279	273
	Fistula (%)	73 (13.2%)	40 (14.3%)	33 (12.1%)
	[95% CI]	[10.4, 16.1]	[10.2, 18.5]	[8.2, 16.0]
	[95% CI]	[0.0, 62.0]	[0.0, 86.7]	[0.0, 67.4]
Surgeon 23	Surgeries	12	6	6
	Fistula (%)	1 (8.3%)	1 (16.7%)	0 (0.0%)
	[95% CI]	[0.0, 24.0]	[0.0, 46.5]	[., .]
Surgeon 24	Surgeries	5	4	1
	Fistula (%)	1 (20.0%)	1 (25.0%)	0 (0.0%)
	[95% CI]	[0.0, 55.1]	[0.0, 67.4]	[., .]
Surgeon 25	Surgeries	9	6	3
	Fistula (%)	1 (11.1%)	1 (16.7%)	0 (0.0%)
	[95% CI]	[0.0, 31.6]	[0.0, 46.5]	[., .]
Surgeon 26	Surgeries	3	2	1
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]

Appendix Material 7: SAS macro to simulate data

```
%macro SIMDATA (RSURG=desired within cluster variance, INTCEPT=original intercept,
TRTEST=desired treatment difference);


DATA SIMDATA;
DO SET=1 TO 2000;
  DO PID=1 TO 521;
  FORMAT SURGEONLAB $266.;
  IF (1<=PID<=69) THEN DO; SURGEONLAB="Surgeon_1"; SURG=1; END;
  IF (69<PID<=80) THEN DO; SURGEONLAB="Surgeon_2"; SURG=2; END;
  IF (80<PID<=117) THEN DO; SURGEONLAB="Surgeon_3"; SURG=3; END;
  IF (117<PID<=158) THEN DO; SURGEONLAB="Surgeon_4"; SURG=4; END;
  IF (158<PID<=180) THEN DO; SURGEONLAB="Surgeon_5"; SURG=5; END;
  IF (180<PID<=184) THEN DO; SURGEONLAB="Surgeon_6"; SURG=6; END;
  IF (184<PID<=203) THEN DO; SURGEONLAB="Surgeon_7"; SURG=7; END;
  IF (203<PID<=214) THEN DO; SURGEONLAB="Surgeon_8"; SURG=8; END;
  IF (214<PID<=235) THEN DO; SURGEONLAB="Surgeon_9"; SURG=9; END;
  IF (235<PID<=320) THEN DO; SURGEONLAB="Surgeon_10"; SURG=10; END;
  IF (320<PID<=326) THEN DO; SURGEONLAB="Surgeon_11"; SURG=11; END;
  IF (326<PID<=348) THEN DO; SURGEONLAB="Surgeon_12"; SURG=12; END;
  IF (348<PID<=368) THEN DO; SURGEONLAB="Surgeon_13"; SURG=13; END;
  IF (368<PID<=374) THEN DO; SURGEONLAB="Surgeon_14"; SURG=14; END;
  IF (374<PID<=379) THEN DO; SURGEONLAB="Surgeon_15"; SURG=15; END;
  IF (379<PID<=393) THEN DO; SURGEONLAB="Surgeon_16"; SURG=16; END;
  IF (393<PID<=418) THEN DO; SURGEONLAB="Surgeon_17"; SURG=17; END;
  IF (418<PID<=438) THEN DO; SURGEONLAB="Surgeon_18"; SURG=18; END;
  IF (438<PID<=460) THEN DO; SURGEONLAB="Surgeon_19"; SURG=19; END;
  IF (460<PID<=468) THEN DO; SURGEONLAB="Surgeon_20"; SURG=20; END;
```

```
 IF (468<PID<=485) THEN DO; SURGEONLAB="Surgeon_21"; SURG=21; END;
 IF (485<PID<=492) THEN DO; SURGEONLAB="Surgeon_22"; SURG=22; END;
 IF (492<PID<=504) THEN DO; SURGEONLAB="Surgeon_23"; SURG=23; END;
 IF (504<PID<=509) THEN DO; SURGEONLAB="Surgeon_24"; SURG=24; END;
 IF (509<PID<=518) THEN DO; SURGEONLAB="Surgeon_25"; SURG=25; END;
 IF (518<PID<=521) THEN DO; SURGEONLAB="Surgeon_26"; SURG=26; END;
 OUTPUT;
 END;
END;
KEEP SET PID SURGEONLAB SURG;
RUN;

/*DEFINE PER LEVEL OF SURGEON*/
DATA RSURG;
CALL streaminit(25345278);
DO SET=1 TO 2000;
 DO SURG=1 TO 26;
 RSURG=rand('normal',0,sqrt(&RSURG));
 OUTPUT;
 END;
END;
RUN;

proc sort data=RSURG out=RSURG;
by SET SURG;
RUN;

proc sort data=SIMDATA out=SIMDATA;
```

```
by SET SURG;
RUN;

DATA SIMDATA;
MERGE SIMDATA RSURG;
BY SET SURG;
RUN;

DATA SIMDATA;
call streaminit(25345278);
SET SIMDATA;
KEEP SET PID SURGEONLAB SURG TRT LOGIT P Y RSURG;
TRT=rand("BERNoulli",0.5); /*DEFINE TRT*/
logit=&INTCEPT+&TRTEST*TRT + RSURG;
p=exp(-logit)/(1+exp(-logit));
if rand('uniform')>p then y=1; else y=0;
RUN;

proc sort data=SIMDATA out=SIMDATA;
by SET SURG PID;
RUN;
%mend SIMDATA;
```

Appendix Material 8: SAS code to apply adjusted model to dataset

```
%macro ORADJ (INDATA=input dataset, OUTDATA=output dataset, GROUP=scenario label, SET=dataset label within scenario);
ods listing close;
ods output OddsRatios=OR;
proc glimmix data=&INDATA;
class SURGEONLAB TRT(ref="0");
model Y = TRT / link=logit dist=binomial s cl alpha=0.05 or;;
random intercept / subject=SURGEONLAB;
nloptions tech=nrridg;
run;
ods output close;
ods listing;
DATA &OUTDATA;
SET OR;
KEEP Estimate Lower Upper GROUP SET MODEL;
GROUP=&GROUP;
SET=&SET;
MODEL="SURGEON ADJUSTED";
RUN;
%mend ORADJ;
%macro repeatORADJ(name);
%do i=1 %to 2000;
%ORADJ(&name..S_&i.,ORAD.&name.S_&i.,"&name","S_&i.");
%end;
%mend;
```

Appendix Material 9: SAS code to apply unadjusted model to dataset

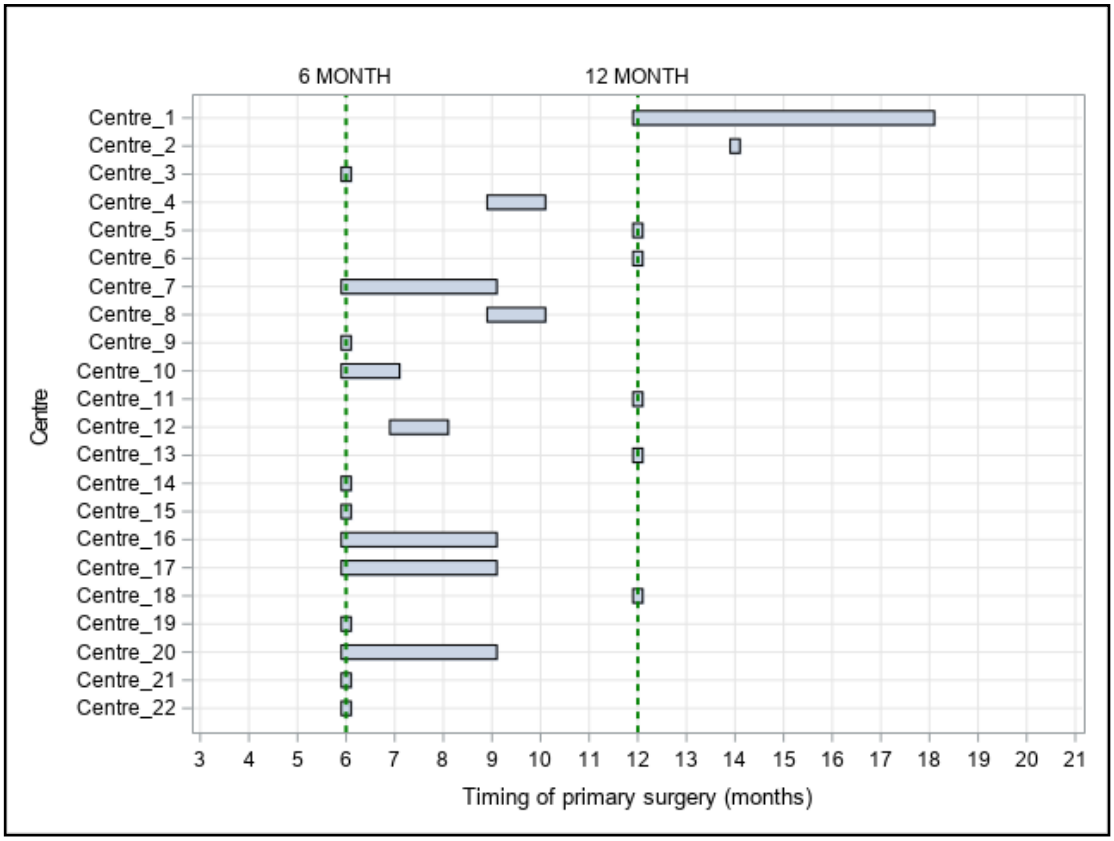
```
%macro ORUNADJ (INDATA=input dataset, OUTDATA=output dataset, GROUP=scenario
label, SET=dataset label within scenario);
ods listing close;
ods output OddsRatios=OR;
proc glimmix data=&INDATA;
class TRT(ref="0");
model Y = TRT / link=logit dist=binomial s cl alpha=0.05 or;;
run;
ods output close;
ods listing;

DATA &OUTDATA;
SET OR;
KEEP Estimate Lower Upper GROUP SET MODEL;
GROUP=&GROUP;
SET=&SET;
MODEL="SURGEON UNADJUSTED";
RUN;
%mend ORUNADJ;

%macro repeatORUN(name);
%do i=1 %to 2000;
%ORUNADJ(&name..S_&i.,ORUN.&name.S_&i.,"&name","S_&i.");
%end;
%mend;
```

Appendix 7 – Supplementary material from Chapter 8

Appendix Figure 1: Timing of surgery for cleft palate at the time of grant application per centre



Appendix Table 10: Summary of prior experience of centres and surgeons

Centre	Prior experience	Technique experience	Surgeon	Sample size	Primary centre	Secondary centre
Centre 1	12-18 months	No	Surgeon 1	69	69	0
			Surgeon 10	85	85	0
Centre 2	14 months	Yes	Surgeon 7	19	18	1
			Surgeon 20	8	7	1
			Surgeon 21	17	17	0
Centre 3	6 months	Yes	Surgeon 4	41	40	1
Centre 4	9-10 months	Yes	Surgeon 3	37	37	0
Centre 5	12 months	Yes	Surgeon 13	20	20	0
			Surgeon 15	5	5	0
Centre 6	12 months	Yes	Surgeon 17	25	25	0
Centre 7	6-9 months	Yes	Surgeon 9	21	21	0
Centre 8	9-10 months	Yes	Surgeon 18	20	20	0
Centre 9	6 months	Yes	Surgeon 12	22	22	0
Centre 10	6-7 months	Yes	Surgeon 5	22	22	0
Centre 11	12 months	Yes	Surgeon 19	22	19	3
Centre 12	7-8 months	Yes	Surgeon 16	14	14	0
Centre 13	12 months	Yes	Surgeon 23	12	12	0
Centre 14	6 months	Yes	Surgeon 2	11	11	0
Centre 15	6 months	Yes	Surgeon 8	11	11	0
Centre 16	6-9 months	Yes	Surgeon 25	9	8	1
Centre 17	6-9 months	Yes	Surgeon 24	5	5	0
Centre 18	12 months	Yes	Surgeon 11	6	6	0
Centre 19	6 months	Yes	Surgeon 14	6	6	0
Centre 20	6-9 months	Yes	Surgeon 22	7	4	3

Appendix Table 10: Summary of prior experience of centres and surgeons

Centre	Prior experience	Technique experience	Surgeon	Sample size	Primary centre	Secondary centre
Centre 21	6 months	Yes	Surgeon 6	4	4	0
Centre 22	6 months	Yes	Surgeon 26	3	3	0

Appendix Table 11: Summary of operation time by surgeon in TOPS

Totals presented based only on infants with operation time data (N=516, see *Table 29*).

		Operation time		
		Overall	Six-months	Twelve-months
Overall	Surgeries	516	264	252
	Mean (SD)	84.7 (37.7)	86.3 (38.2)	82.9 (35.7)
	[Min, Max]	[30.0, 245.0]	[30.0, 245.0]	[30.0, 210.0]
Surgeon 1	Surgeries	69	30	39
	Mean (SD)	133.6 (33.9)	143.5 (36.2)	125.9 (30.0)
	[Min, Max]	[80, 245]	[82, 245]	[80, 210]
Surgeon 2	Surgeries	11	5	6
	Mean (SD)	147.5 (37.2)	148.0 (44.4)	147.0 (34.5)
	[Min, Max]	[109, 213]	[109, 213]	[112, 200]
Surgeon 3	Surgeries	37	19	18
	Mean (SD)	68.7 (11.1)	70.3 (11.0)	67.1 (11.3)
	[Min, Max]	[50, 90]	[52, 90]	[50, 90]
Surgeon 4	Surgeries	41	22	19
	Mean (SD)	52.2 (9.6)	52.9 (9.3)	51.4 (10.2)
	[Min, Max]	[37, 75]	[37, 70]	[38, 75]
Surgeon 5	Surgeries	22	12	10
	Mean (SD)	49.8 (11.6)	53.3 (13.0)	45.6 (8.4)
	[Min, Max]	[30, 75]	[36, 75]	[30, 55]
Surgeon 6	Surgeries	4	1	3
	Mean (SD)	70.5 (17.5)	53 (.)	76.3 (16.0)
	[Min, Max]	[53, 92]	[53, 53]	[60, 92]
Surgeon 7	Surgeries	18	9	9
	Mean (SD)	74.3 (12.4)	74.1 (13.4)	74.6 (12.1)

Appendix Table 11: Summary of operation time by surgeon in TOPS

Totals presented based only on infants with operation time data (N=516, see *Table 29*).

		Operation time		
		Overall	Six-months	Twelve-months
Overall	Surgeries	516	264	252
	Mean (SD)	84.7 (37.7)	86.3 (38.2)	82.9 (35.7)
	[Min, Max]	[30.0, 245.0]	[30.0, 245.0]	[30.0, 210.0]
	[Min, Max]	[56, 99]	[57, 99]	[56, 91]
Surgeon 8	Surgeries	11	6	5
	Mean (SD)	91.6 (16.4)	98 (17.5)	84.0 (12.4)
	[Min, Max]	[65, 125]	[75, 125]	[65, 95]
Surgeon 9	Surgeries	21	13	8
	Mean (SD)	60.3 (15.3)	65.8 (13.9)	51.3 (13.6)
	[Min, Max]	[30, 90]	[50, 90.0]	[30, 65]
Surgeon 10	Surgeries	85	44	41
	Mean (SD)	63.3 (29.1)	69.4 (33.8)	56.7 (21.5)
	[Min, Max]	[32, 165]	[32, 165]	[34, 137]
Surgeon 11	Surgeries	6	2	4
	Mean (SD)	102.8 (23.2)	107.0 (24.0)	100.8 (24.2)
	[Min, Max]	[80, 139]	[90, 124]	[80, 139]
Surgeon 12	Surgeries	22	11	11
	Mean (SD)	103.8 (24.6)	103.2 (17.4)	104.4 (31.1)
	[Min, Max]	[65, 165]	[75, 120]	[65, 165]
Surgeon 13	Surgeries	20	12	8
	Mean (SD)	64.2 (12.8)	65.8 (10.0)	61.8 (16.7)
	[Min, Max]	[49, 101]	[54, 83]	[49, 101]
Surgeon 14	Surgeries	6	3	3

Appendix Table 11: Summary of operation time by surgeon in TOPS

Totals presented based only on infants with operation time data (N=516, see *Table 29*).

		Operation time		
		Overall	Six-months	Twelve-months
Overall	Surgeries	516	264	252
	Mean (SD)	84.7 (37.7)	86.3 (38.2)	82.9 (35.7)
	[Min, Max]	[30.0, 245.0]	[30.0, 245.0]	[30.0, 210.0]
	Mean (SD)	93.3 (18.4)	93.3 (6.5)	93.3 (28.4)
	[Min, Max]	[68, 124]	[87, 100]	[68, 124]
Surgeon 15	Surgeries	5	1	4
	Mean (SD)	111.4 (23.0)	147.0 (.)	102.5 (13.3)
	[Min, Max]	[90, 147]	[147, 147]	[90, 121]
Surgeon 16	Surgeries	14	6	8
	Mean (SD)	87.7 (24.0)	86.7 (33.5)	88.5 (16.2)
	[Min, Max]	[43, 135]	[43, 135]	[65, 105]
Surgeon 17	Surgeries	25	12	13
	Mean (SD)	66.1 (23.0)	69.6 (24.1)	62.8 (22.4)
	[Min, Max]	[30, 129]	[43, 129]	[30, 112]
Surgeon 18	Surgeries	19	10	9
	Mean (SD)	98.8 (24.2)	104.3 (26.8)	92.8 (20.7)
	[Min, Max]	[60, 165]	[84, 165]	[60, 135]
Surgeon 19	Surgeries	22	12	10
	Mean (SD)	125.8 (24.7)	128.3 (26.1)	122.7 (24.0)
	[Min, Max]	[95, 180]	[95, 180]	[98, 180]
Surgeon 20	Surgeries	8	5	3
	Mean (SD)	99.0 (37.4)	90.6 (43.5)	113 (25.5)
	[Min, Max]	[52, 165]	[52, 165]	[88, 139]

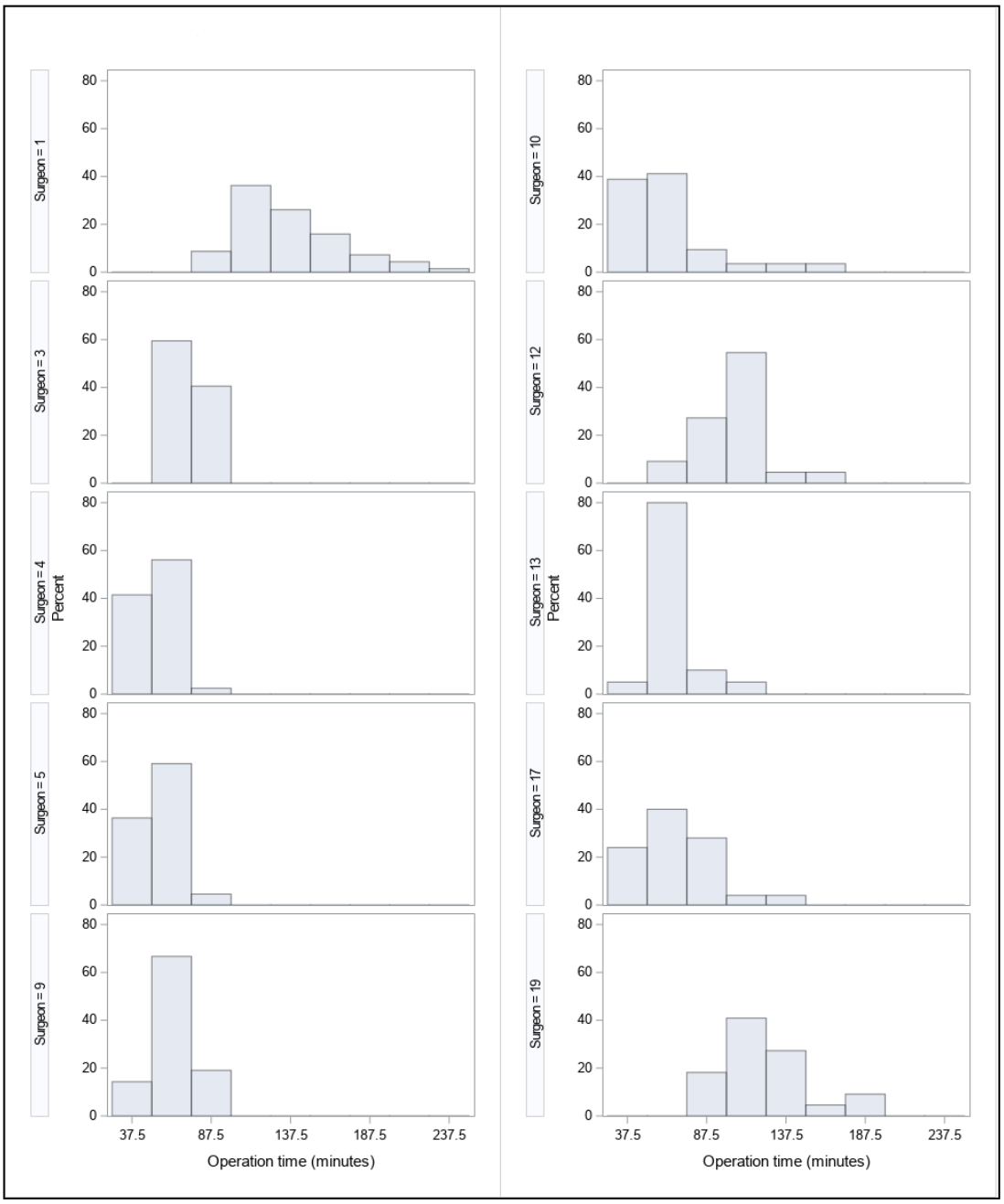
Appendix Table 11: Summary of operation time by surgeon in TOPS

Totals presented based only on infants with operation time data (N=516, see *Table 29*).

		Operation time		
		Overall	Six-months	Twelve-months
Overall	Surgeries	516	264	252
	Mean (SD)	84.7 (37.7)	86.3 (38.2)	82.9 (35.7)
	[Min, Max]	[30.0, 245.0]	[30.0, 245.0]	[30.0, 210.0]
Surgeon 21	Surgeries	15	8	7
	Mean (SD)	108.3 (17.1)	109.4 (16.0)	107.0 (19.5)
	[Min, Max]	[85, 149]	[85, 136]	[93, 149]
Surgeon 22	Surgeries	6	3	3
	Mean (SD)	51.5 (18.0)	44.7 (12.9)	58.3 (22.5)
	[Min, Max]	[30, 80]	[30, 54]	[35, 80]
Surgeon 23	Surgeries	12	6	6
	Mean (SD)	81.5 (11.8)	86.0 (13.9)	77.0 (8.2)
	[Min, Max]	[60, 100]	[60, 100]	[67, 89]
Surgeon 24	Surgeries	5	4	1
	Mean (SD)	59.0 (9.8)	59.0 (11.3)	59.0 (.)
	[Min, Max]	[45, 70]	[45, 70]	[59, 59]
Surgeon 25	Surgeries	9	6	3
	Mean (SD)	78.9 (16.8)	81.7 (20.0)	73.3 (7.6)
	[Min, Max]	[63, 105]	[63, 105]	[65, 80]
Surgeon 26	Surgeries	3	2	1
	Mean (SD)	91.3 (4.0)	91.0 (5.7)	92 (.)
	[Min, Max]	[87, 95]	[87, 95]	[92, 92]

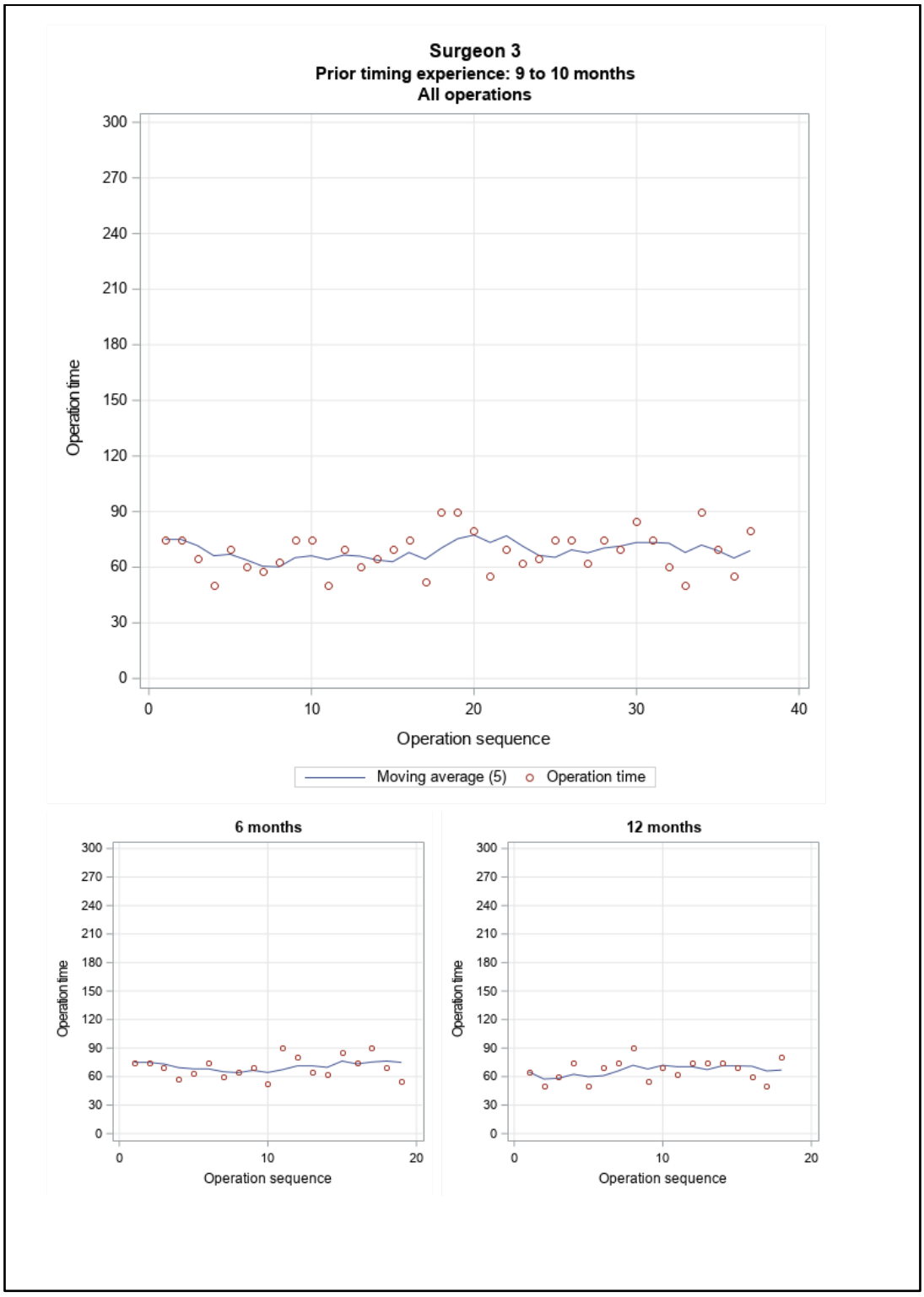
Appendix Figure 2: Histogram of operation times by surgeon

Figures presented for surgeons who operated on at least twenty patients only.



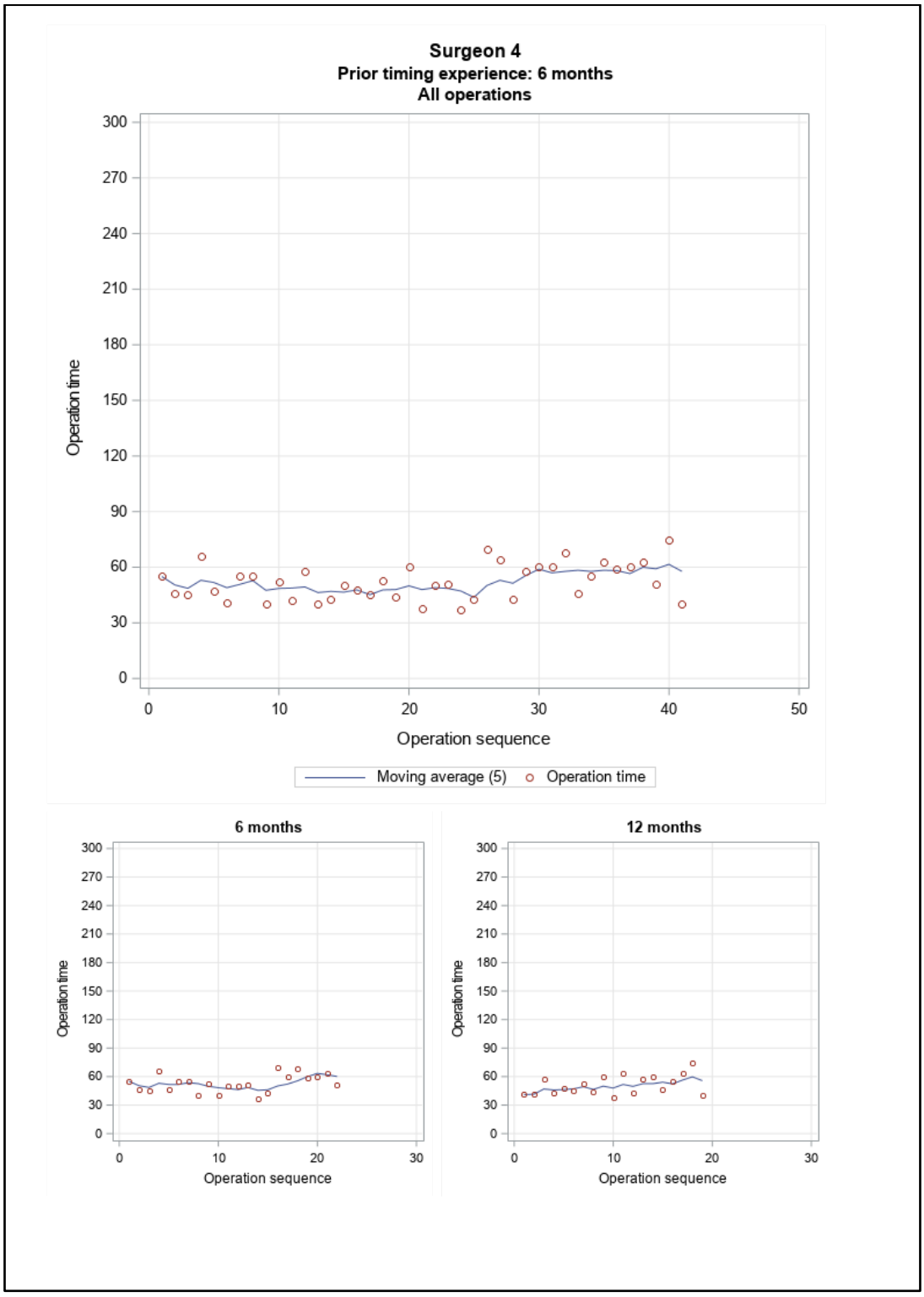
Appendix Figure 3: Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



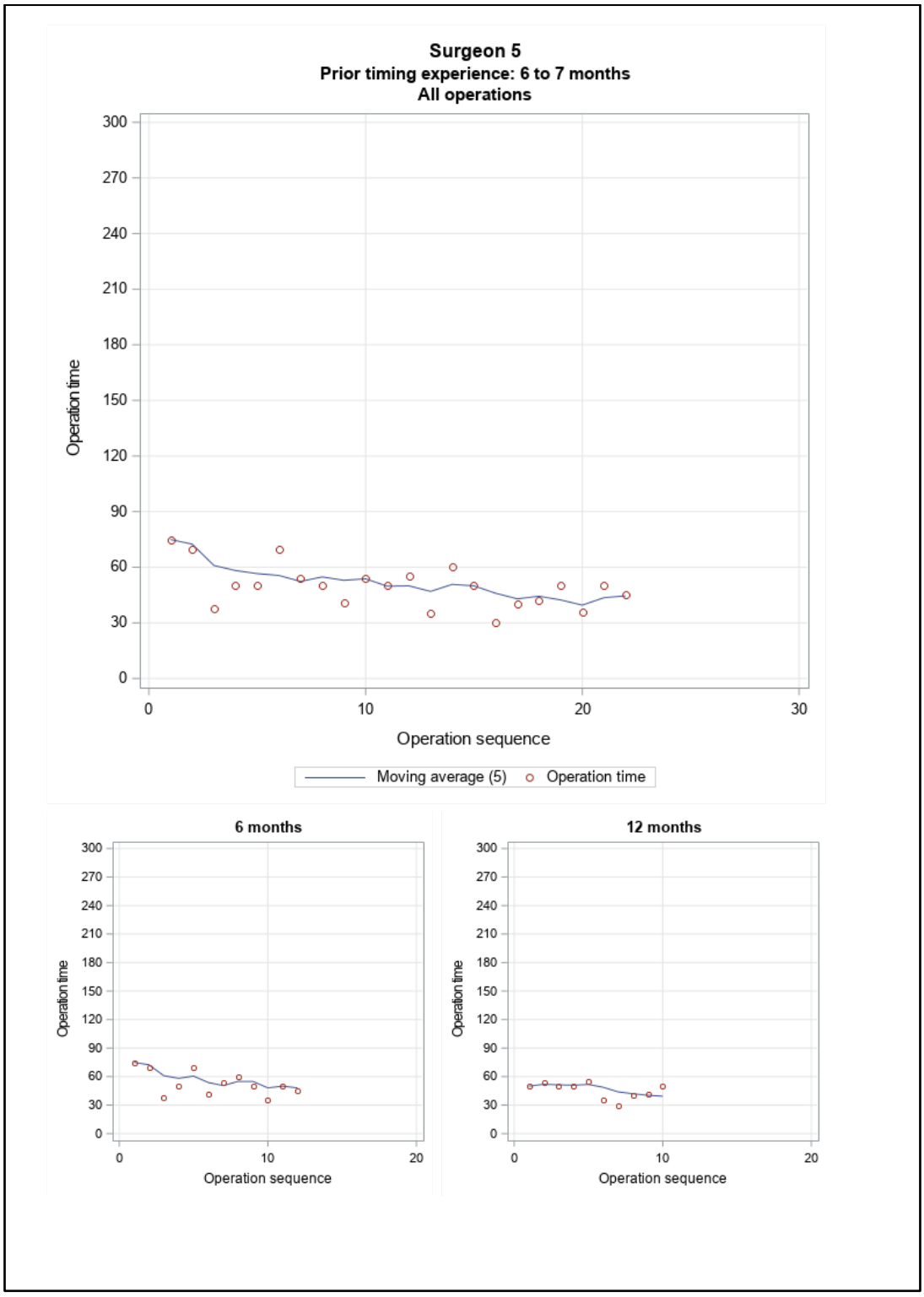
Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



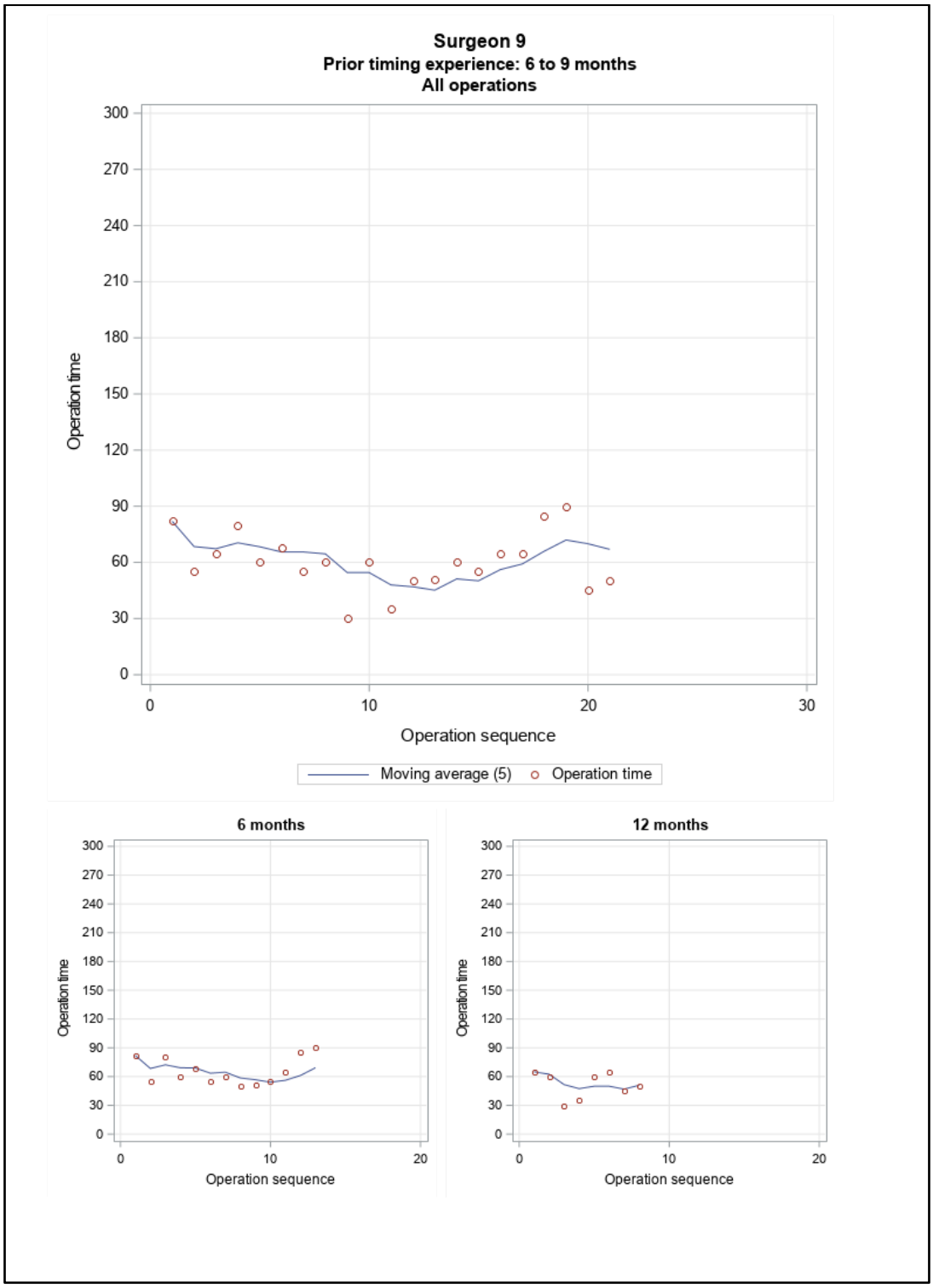
Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



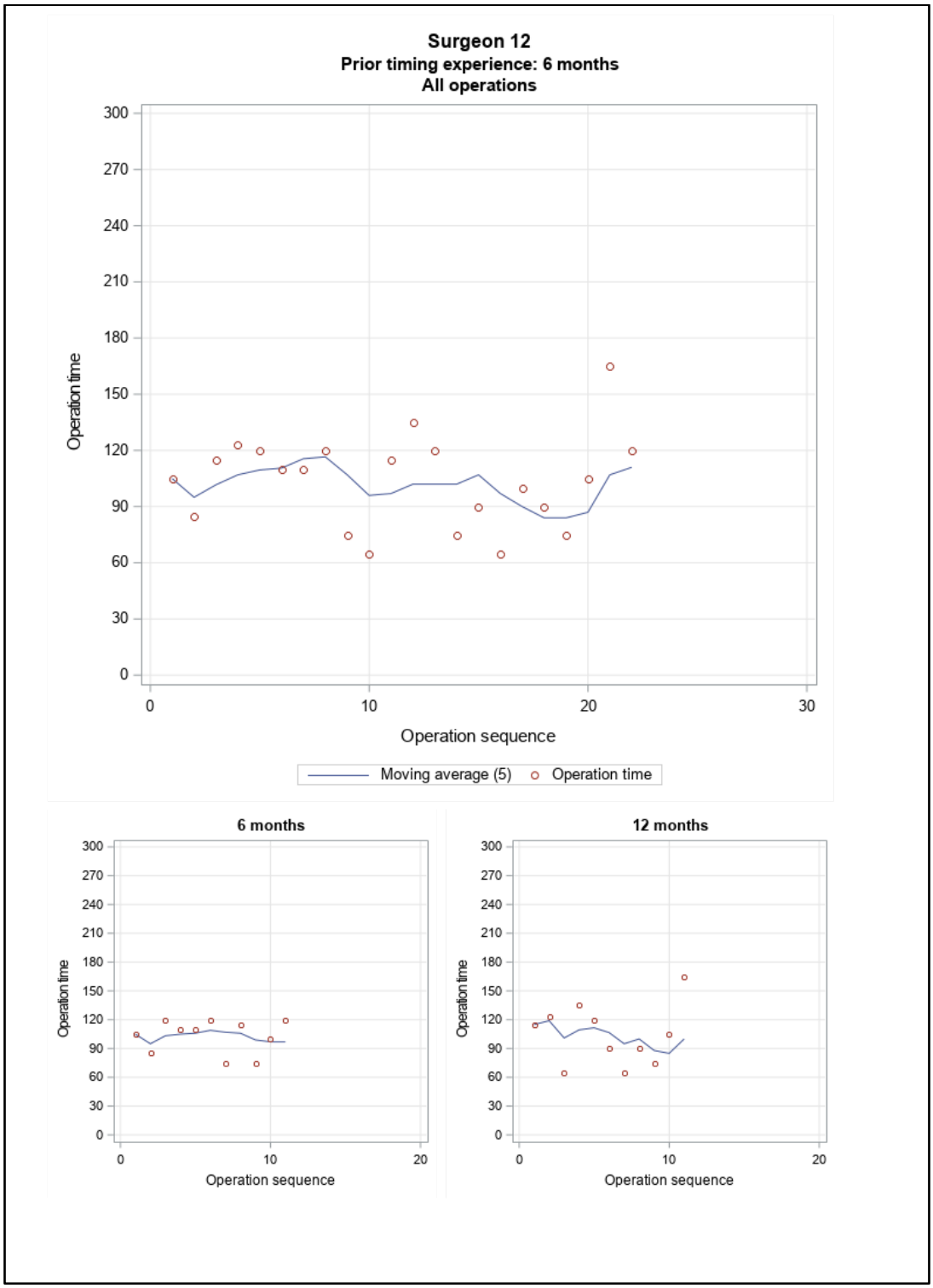
Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



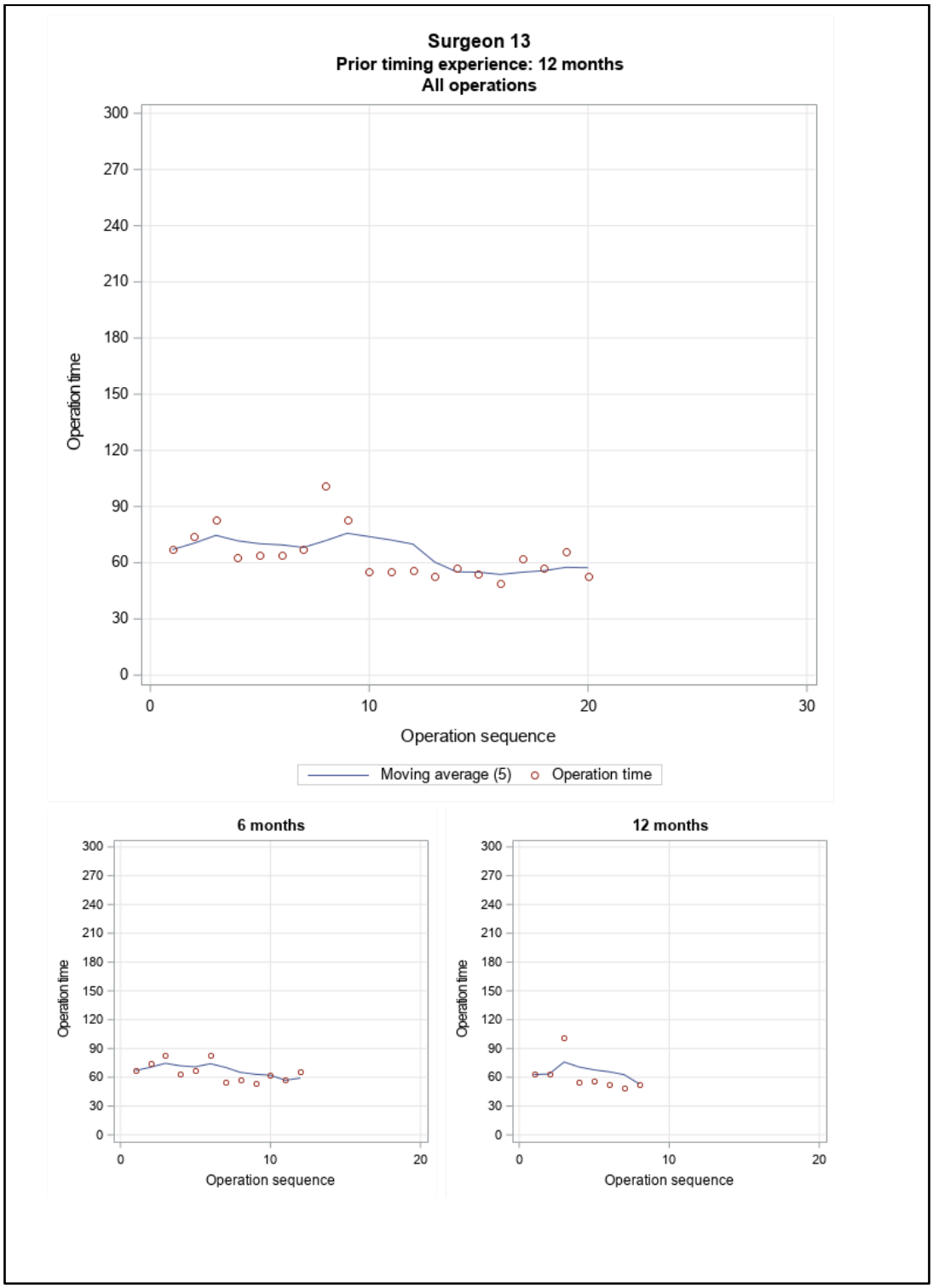
Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



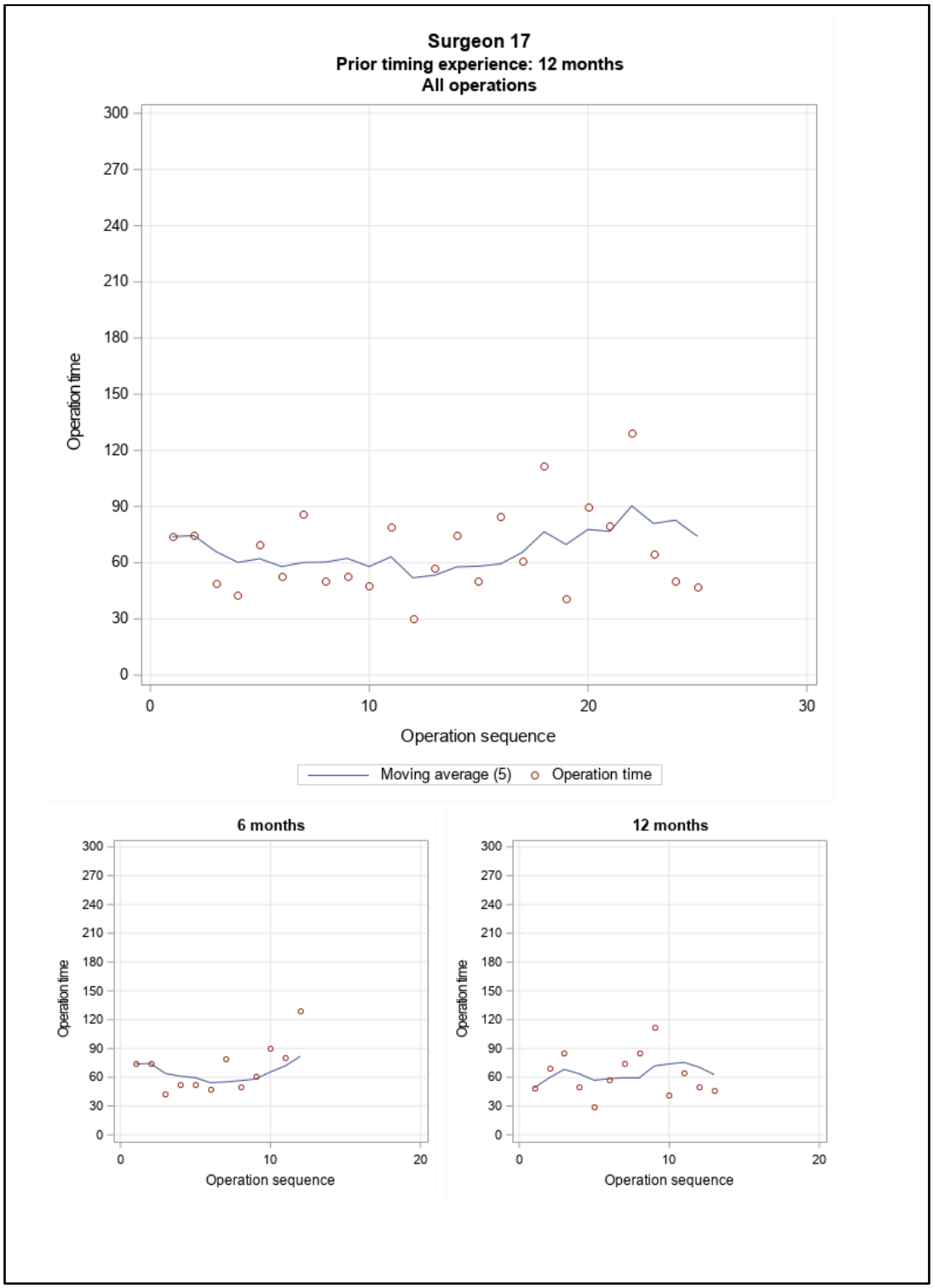
Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



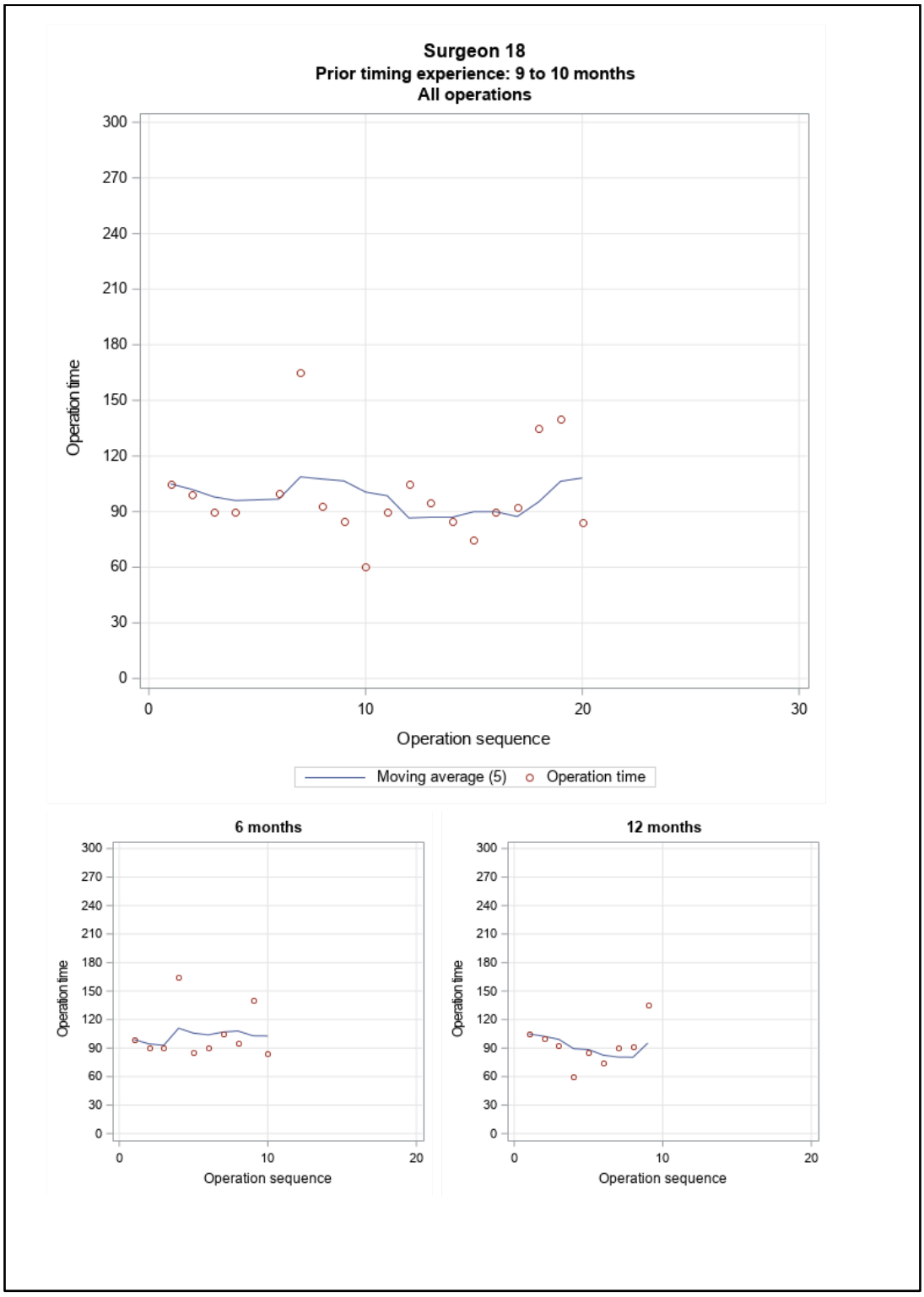
Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



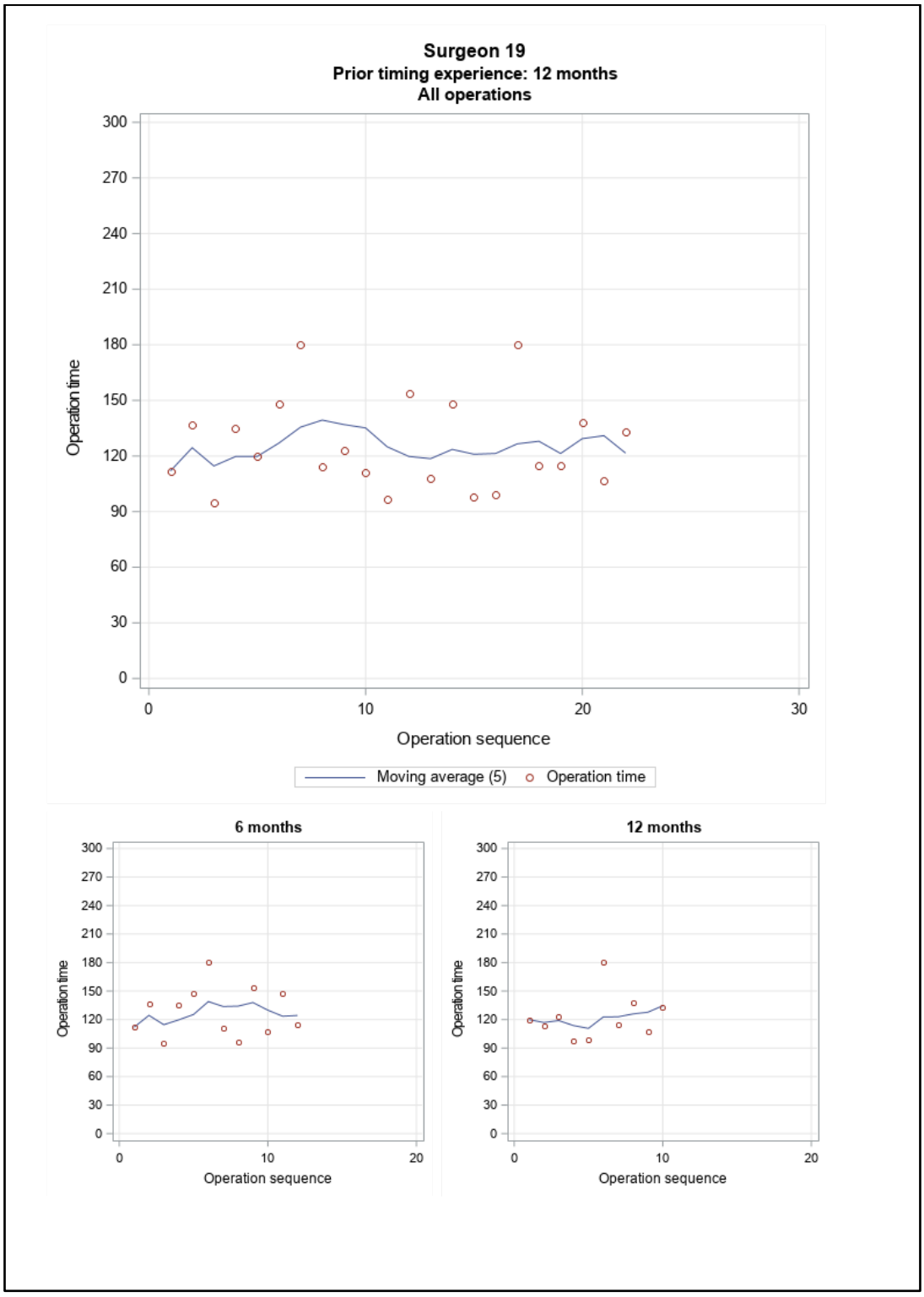
Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



Appendix Figure 3 (continued): Moving average operation time against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. Moving averages of order five, based on an equal weight distribution, presented.



Appendix Table 12: Summary of fistula by surgeon in TOPS

Totals presented based only on infants with surgery data (N=521, see *Table 29*).

		Fistula		
		Overall	Six-months	Twelve-months
Overall	Surgeries	521	266	255
	Fistula (%)	73 (14.0%)	40 (15.0%)	33 (12.9%)
	[95% CI]	[11.0, 17.0]	[10.7, 19.3]	[8.8, 17.1]
Surgeon 1	Surgeries	69	30	39
	Fistula (%)	9 (13.0%)	4 (13.3%)	5 (12.8%)
	[95% CI]	[5.1, 21.0]	[1.2, 25.5]	[2.3, 23.3]
Surgeon 2	Surgeries	11	5	6
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 3	Surgeries	37	19	18
	Fistula (%)	7 (18.9%)	4 (21.1%)	3 (16.7%)
	[95% CI]	[6.3, 31.5]	[2.7, 39.4]	[0.0, 33.9]
Surgeon 4	Surgeries	41	22	19
	Fistula (%)	8 (19.5%)	5 (22.7%)	3 (15.8%)
	[95% CI]	[7.4. 31.6]	[5.2, 40.2]	[0.0, 32.2]
Surgeon 5	Surgeries	22	12	10
	Fistula (%)	2 (9.1%)	2 (16.7%)	0 (0.0%)
	[95% CI]	[0.0, 21.1]	[0.0, 37.8]	[., .]
Surgeon 6	Surgeries	4	1	3
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 7	Surgeries	19	10	9
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Appendix Table 12: Summary of fistula by surgeon in TOPS

Totals presented based only on infants with surgery data (N=521, see *Table 29*).

		Fistula		
		Overall	Six-months	Twelve-months
Overall	Surgeries	521	266	255
	Fistula (%)	73 (14.0%)	40 (15.0%)	33 (12.9%)
	[95% CI]	[11.0, 17.0]	[10.7, 19.3]	[8.8, 17.1]
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 8	Surgeries	11	6	5
	Fistula (%)	3 (27.3%)	3 (50.0%)	0 (0.0%)
	[95% CI]	[1.0, 53.6]	[10.0, 90.0]	[., .]
Surgeon 9	Surgeries	21	13	8
	Fistula (%)	2 (9.5%)	2 (15.4%)	0 (0.0%)
	[95% CI]	[0.0, 22.1]	[0.0, 35.0]	[., .]
Surgeon 10	Surgeries	85	44	41
	Fistula (%)	8 (9.4%)	4 (9.1%)	4 (9.8%)
	[95% CI]	[3.2, 15.6]	[0.6, 17.6]	[0.7, 18.8]
Surgeon 11	Surgeries	6	2	4
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 12	Surgeries	22	11	11
	Fistula (%)	4 (18.2%)	1 (9.1%)	3 (27.3%)
	[95% CI]	[2.1, 34.3]	[0.0, 26.1]	[1.0, 53.6]
Surgeon 13	Surgeries	20	12	8
	Fistula (%)	2 (10.0%)	1 (8.3%)	1 (12.5%)
	[95% CI]	[0.0, 23.1]	[0.0, 24.0]	[0.0, 35.4]
Surgeon 14	Surgeries	6	3	3

Appendix Table 12: Summary of fistula by surgeon in TOPS

Totals presented based only on infants with surgery data (N=521, see *Table 29*).

		Fistula		
		Overall	Six-months	Twelve-months
Overall	Surgeries	521	266	255
	Fistula (%)	73 (14.0%)	40 (15.0%)	33 (12.9%)
	[95% CI]	[11.0, 17.0]	[10.7, 19.3]	[8.8, 17.1]
	Fistula (%)	3 (50.0%)	1 (33.3%)	2 (66.7%)
	[95% CI]	[10.0, 90.0]	[0.0, 86.7]	[13.3, 100.0]
Surgeon 15	Surgeries	5	1	4
	Fistula (%)	1 (20.0%)	0 (0.0%)	1 (25.0%)
	[95% CI]	[0.0, 55.1]	[., .]	[0.0, 67.4]
Surgeon 16	Surgeries	14	6	8
	Fistula (%)	3 (21.4%)	2 (33.3%)	1 (12.5%)
	[95% CI]	[0.0, 42.9]	[0.0, 71.1]	[0.0, 35.4]
Surgeon 17	Surgeries	25	12	13
	Fistula (%)	2 (8.0%)	1 (8.3%)	1 (7.7%)
	[95% CI]	[0.0, 18.6]	[0.0, 24.0]	[0.0, 22.2]
Surgeon 18	Surgeries	20	10	10
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]
Surgeon 19	Surgeries	22	12	10
	Fistula (%)	6 (27.3%)	2 (16.7%)	4 (40.0%)
	[95% CI]	[8.7, 45.9]	[0.0, 37.8]	[9.6, 70.4]
Surgeon 20	Surgeries	8	5	3
	Fistula (%)	3 (37.5%)	2 (40.0%)	1 (33.3%)
	[95% CI]	[4.0, 71.0]	[0.0, 82.9]	[0.0, 86.7]

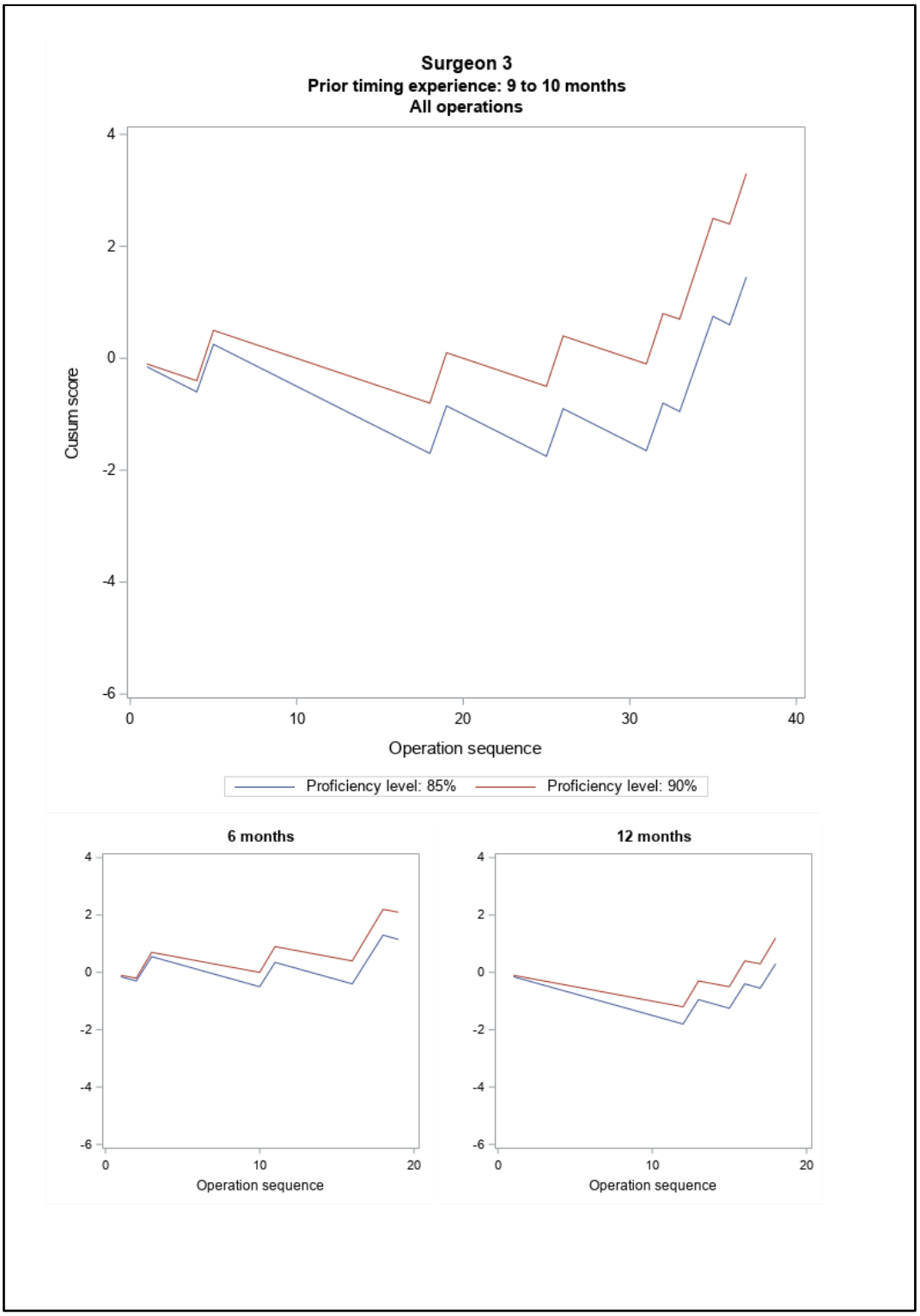
Appendix Table 12: Summary of fistula by surgeon in TOPS

Totals presented based only on infants with surgery data (N=521, see *Table 29*).

		Fistula		
		Overall	Six-months	Twelve-months
Overall	Surgeries	521	266	255
	Fistula (%)	73 (14.0%)	40 (15.0%)	33 (12.9%)
	[95% CI]	[11.0, 17.0]	[10.7, 19.3]	[8.8, 17.1]
Surgeon 21	Surgeries	17	9	8
	Fistula (%)	5 (29.4%)	2 (22.2%)	3 (37.5%)
	[95% CI]	[7.8, 51.1]	[0.0, 49.4]	[4.0, 71.0]
Surgeon 22	Surgeries	7	3	4
	Fistula (%)	2 (28.6%)	1 (33.3%)	1 (25.0%)
	[95% CI]	[0.0, 62.0]	[0.0, 86.7]	[0.0, 67.4]
Surgeon 23	Surgeries	12	6	6
	Fistula (%)	1 (8.3%)	1 (16.7%)	0 (0.0%)
	[95% CI]	[0.0, 24.0]	[0.0, 46.5]	[., .]
Surgeon 24	Surgeries	5	4	1
	Fistula (%)	1 (20.0%)	1 (25.0%)	0 (0.0%)
	[95% CI]	[0.0, 55.1]	[0.0, 67.4]	[., .]
Surgeon 25	Surgeries	9	6	3
	Fistula (%)	1 (11.1%)	1 (16.7%)	0 (0.0%)
	[95% CI]	[0.0, 31.6]	[0.0, 46.5]	[., .]
Surgeon 26	Surgeries	3	2	1
	Fistula (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	[95% CI]	[., .]	[., .]	[., .]

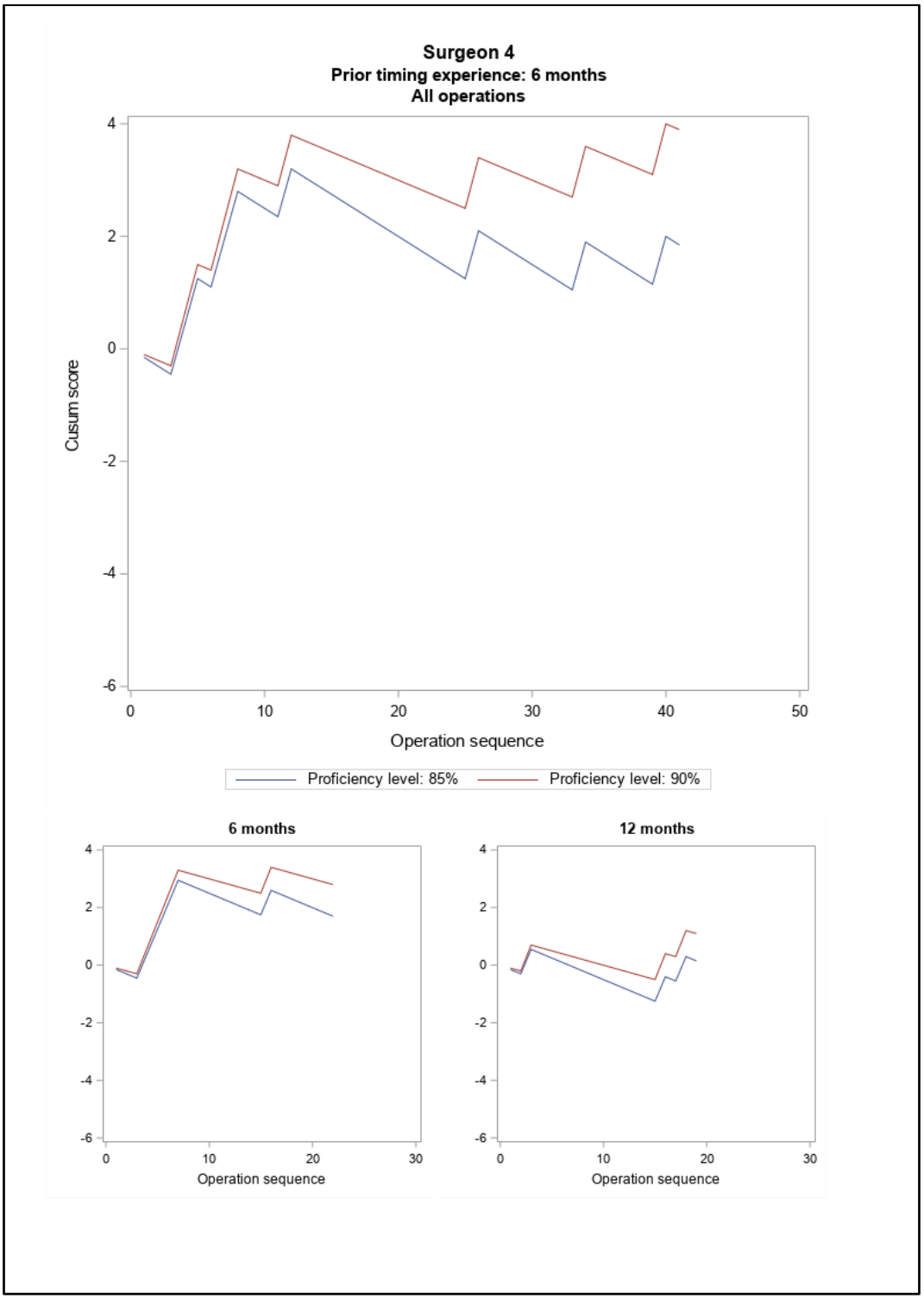
Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



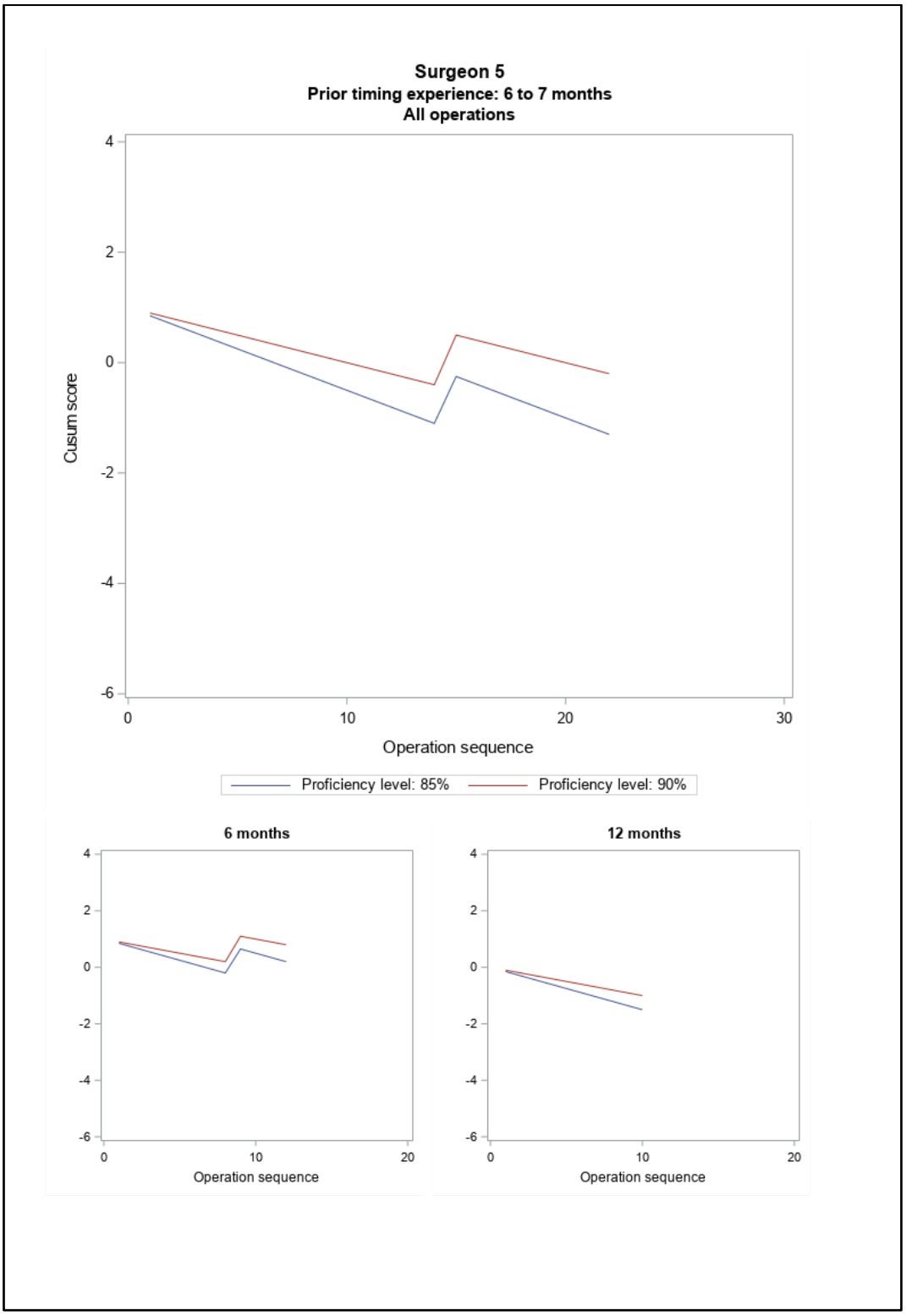
Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



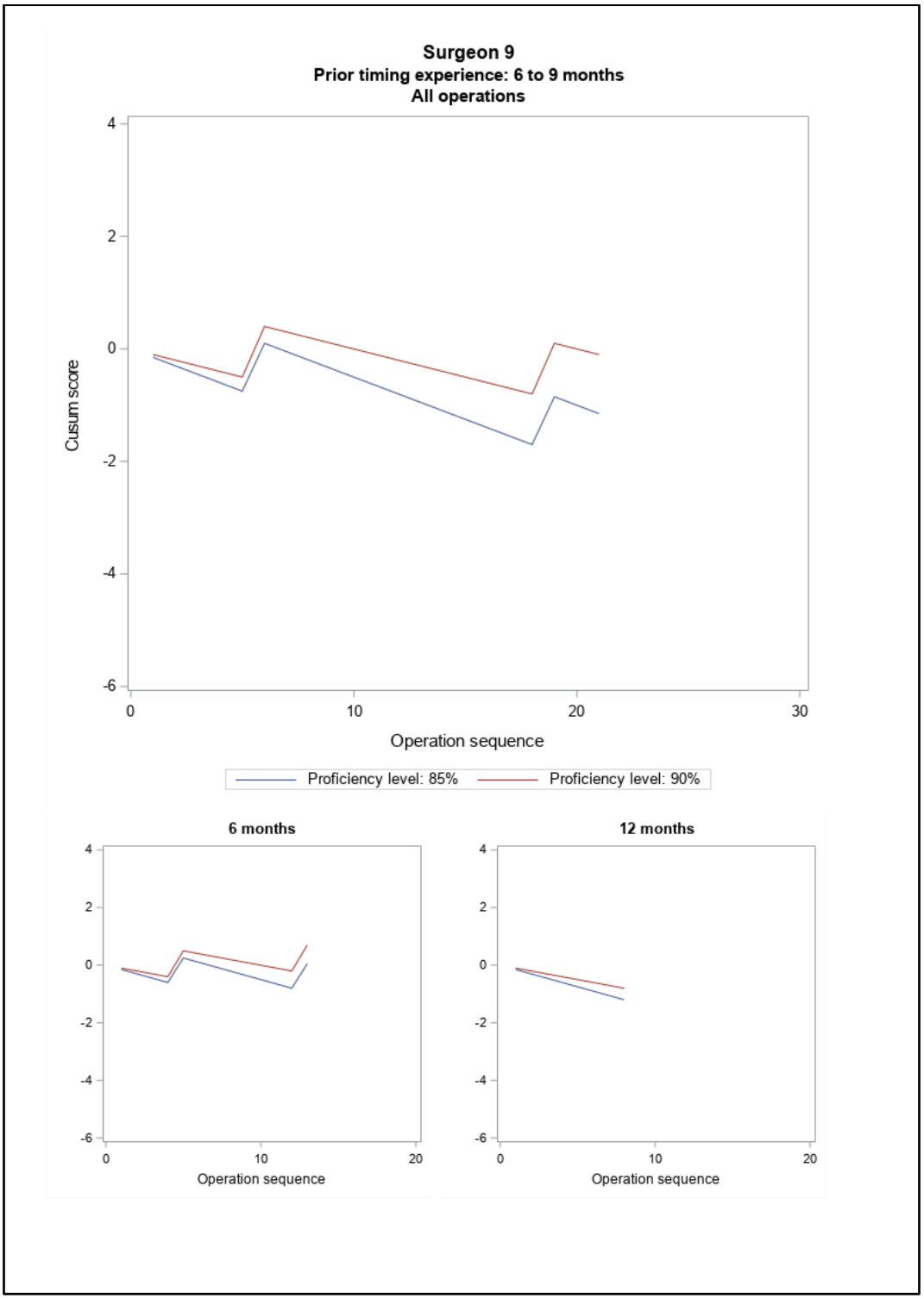
Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



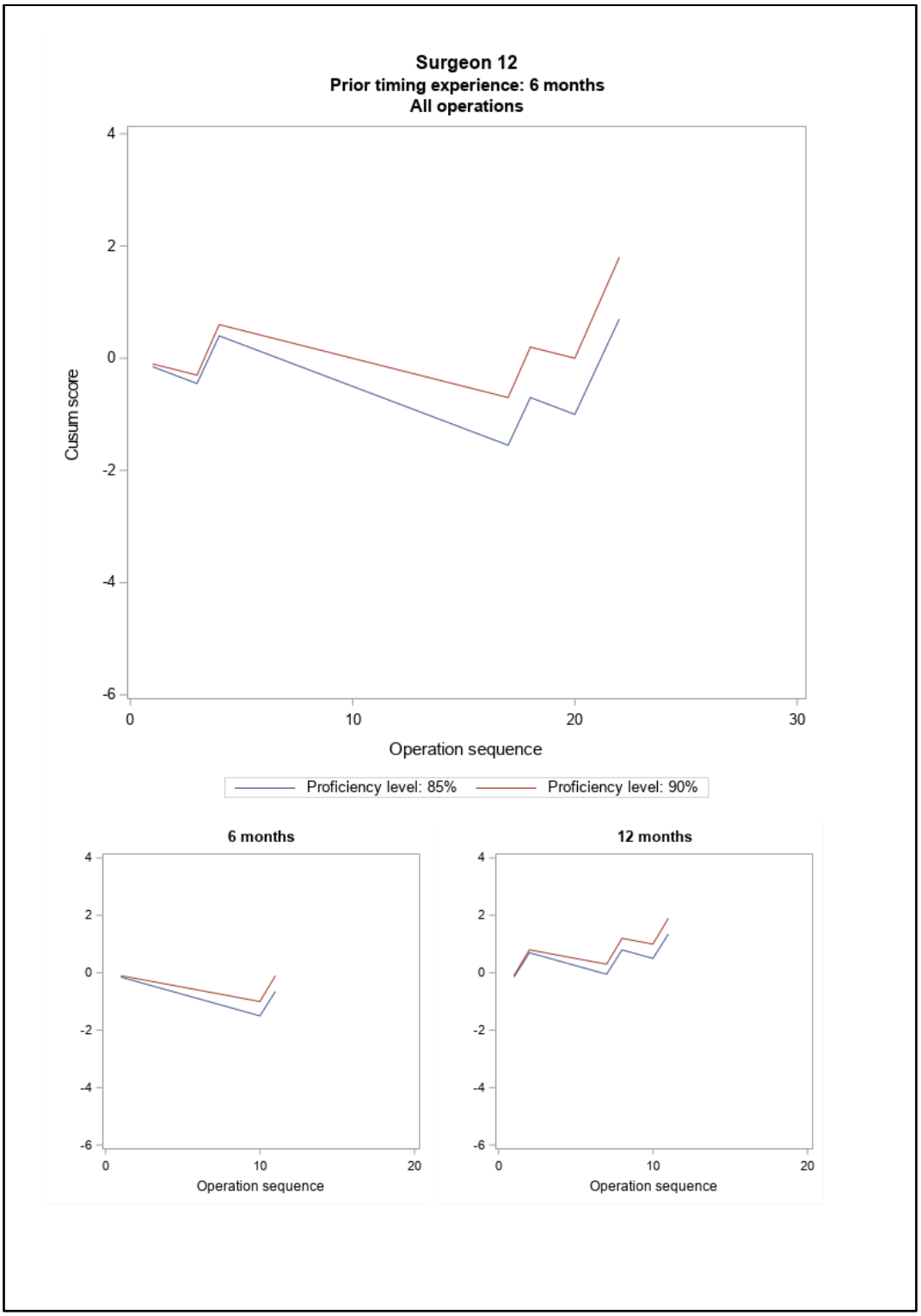
Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



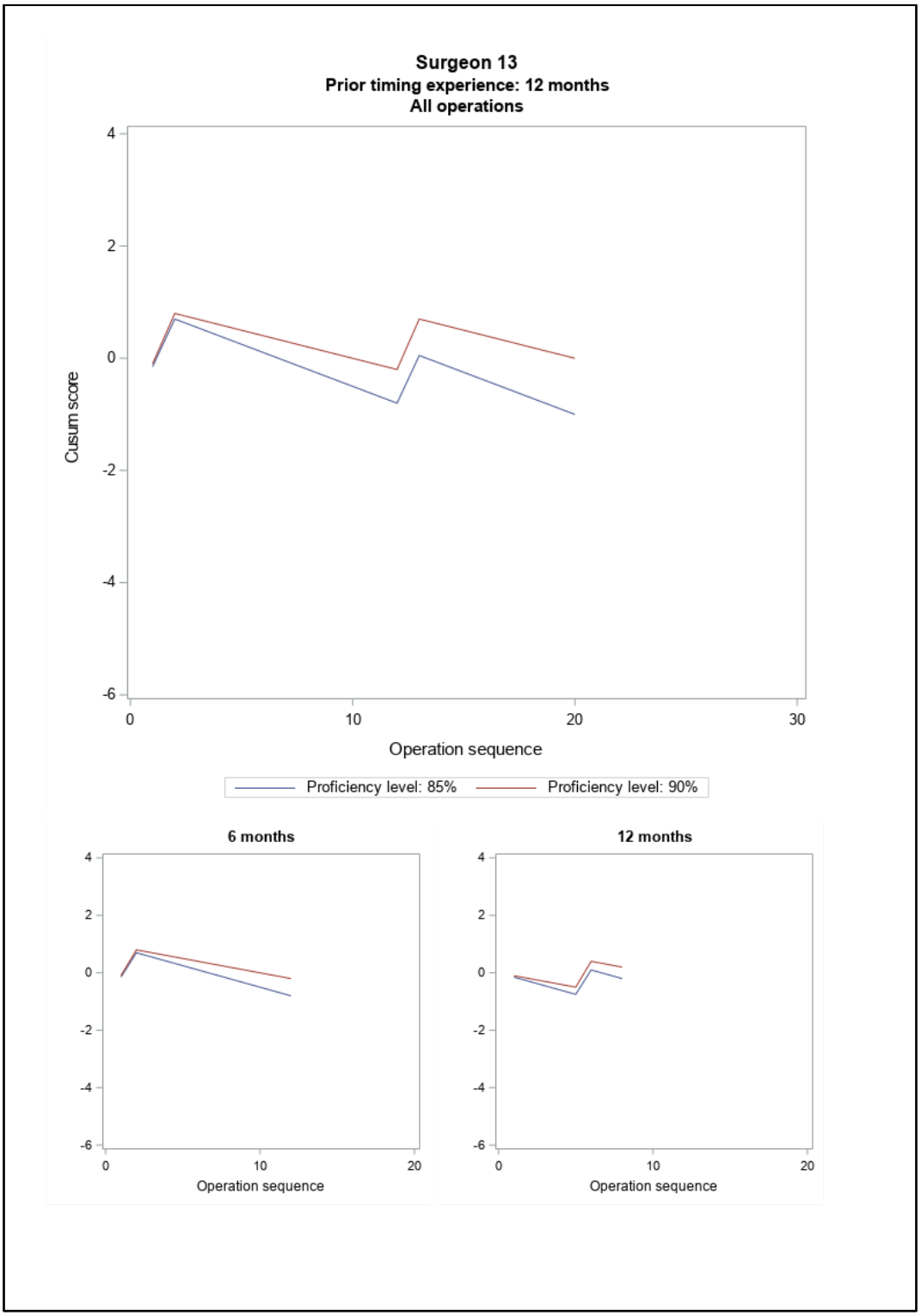
Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



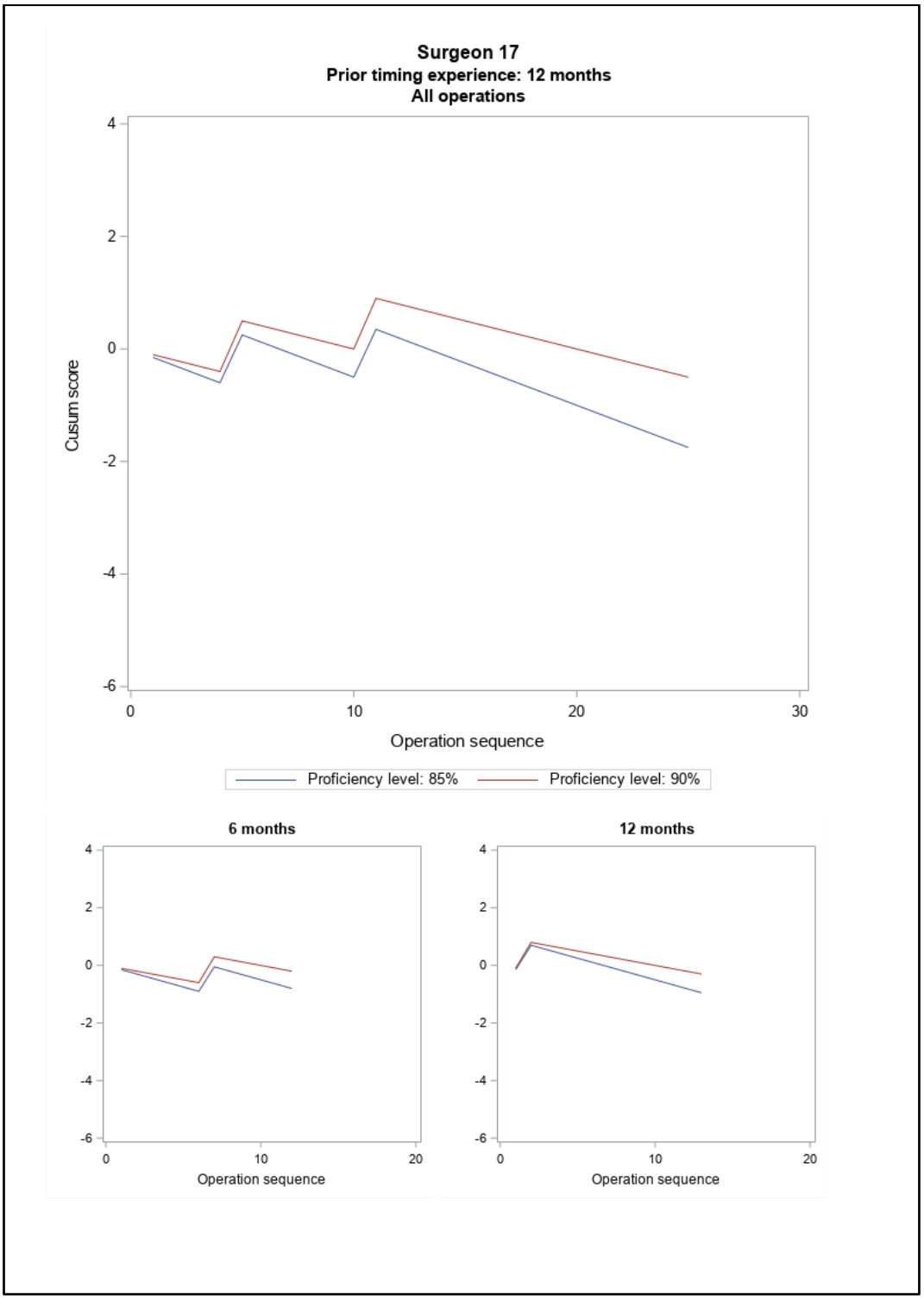
Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



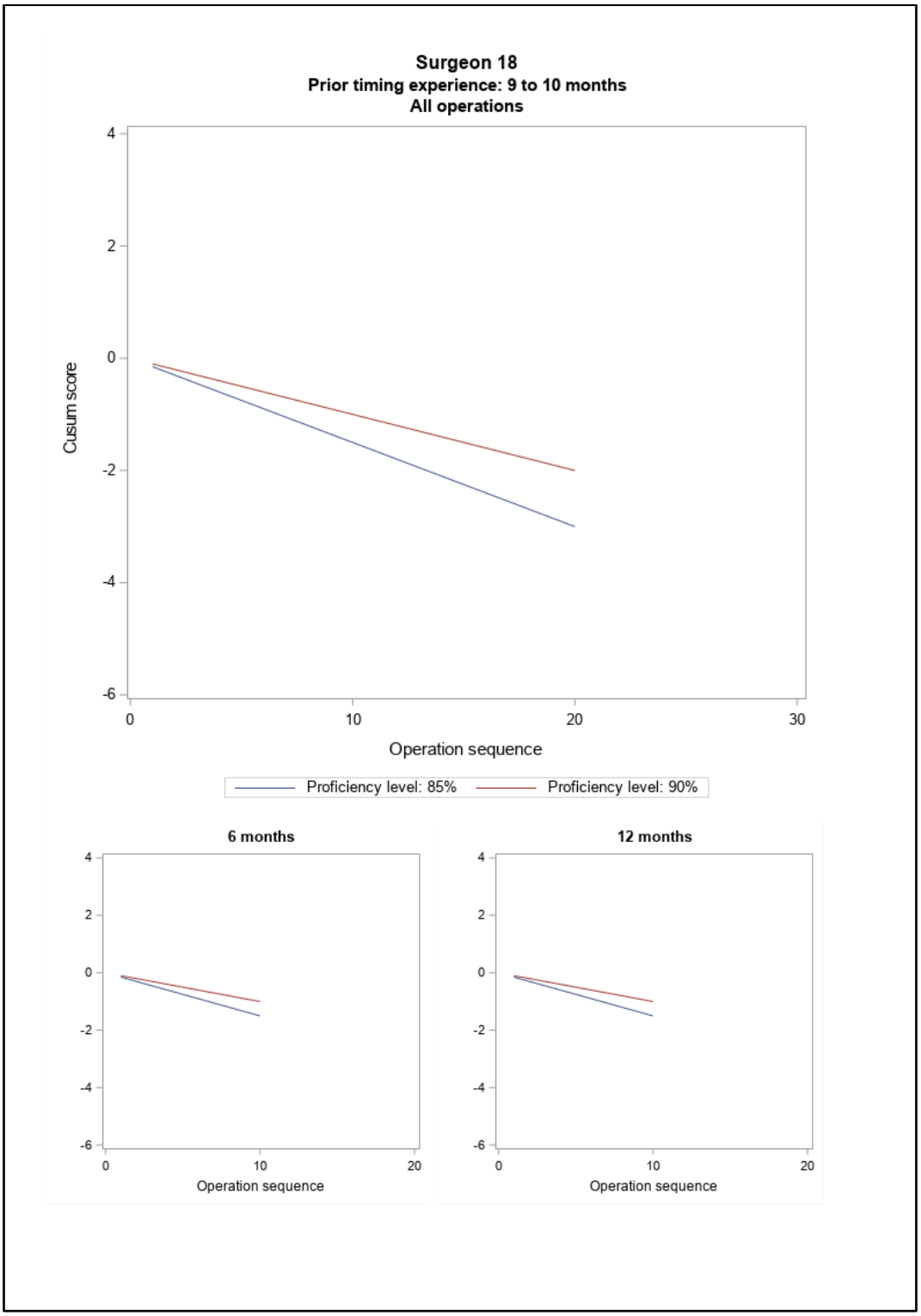
Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.



Appendix Figure 4: Cusum charts for occurrence of fistula against operation sequence

Figures presented for surgeons who operated on at least twenty patients only. The y-axis of each figure represents the cusum score as defined in *Equation 5*.