OPEN ACCESS

University of BRISTOL

Metcalfe, C., & Barham, C. P. (2019, Sep 12). The ROMIO Study: Statistical Analysis Plan. *Bristol: University of Bristol*.

Publisher's PDF, also known as Version of record

License (if available):
Unspecified

PDF-document

Internal document pre-specifying the main analysis for the ROMIO trial - to be made freely available please.

University of Bristol - Explore Bristol Research
General rights

University of
BRISTOL

CTEU Bristol

Bristol Randomised Trials Collaboration

The ROM₿O Study

Statistical Analysis Plan

	NAME	TITLE	SIGNATURE	DATE
Author	Chris Metcalfe	Professor of Medical Statistics, University of Bristol.	*Chris Metcalfe*	13/09/2019
Study CI	Paul Barham	Consultant Surgeon, Bristol Royal Infirmary.	*Paul Barham*	13/09/2019

Table of Contents

Abbreviations

DMSC	Data Monitoring and Safety Committee
EORTC	European Organisation for Research and Treatment of Cancer
FEV$_1$	Forced Expiratory Volume in one second (lung function)
FVC	Forced Vital Capacity (lung function)
HDU	High Dependency Unit
ITU	Intensive Treatment Unit
LAO	Laparoscopically Assisted Oesophagectomy
MDT	Multi-Disciplinary Team
NIHR	National Institute of Health Research
OO	Open Oesophagectomy
PRO	Patient Reported Outcome
ROMIO	Randomised Open or Minimally Invasive Oesophagectomy
SD	Standard Deviation
SMG	Study Management Group
SSC	Study Steering Committee
TMIO	Totally Minimally Invasive Oesophagectomy

1. INTRODUCTION & PURPOSE

This document details the rules proposed and the presentation that will be followed, as closely as possible, when analysing and reporting the main results from the ROMIO study.

The purpose of the plan is to:
1. Ensure that the analysis is appropriate for the aims of the trial, reflects good statistical practice, and that interpretation of a priori and post hoc analyses respectively is appropriate.
2. Explain in detail how the data will be handled and analyzed to enable others to perform the actual analysis in the event of sickness or other absence

Additional exploratory or auxiliary analyses of data not specified in the protocol are permitted but fall outside the scope of this analysis plan (although such analyses would be expected to follow Good Statistical Practice).

The analysis strategy will be made available if required by journal editors or referees when the main papers are submitted for publication. Additional analyses suggested by investigators, reviewers or editors will, if considered appropriate, be performed in accordance with the Analysis Plan, but if reported the source of such a post-hoc analysis will be declared.

Revisions of the statistical analysis plan will include a table of changes subsequent to version 1.0.

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

2. SYNOPSIS OF STUDY DESIGN AND PROCEDURES
2.1 Trial objectives and aims

The following study synopsis has been written with the sole purpose of informing this statistical analysis plan. For details of the study to inform any other purpose, the current version of the protocol must be consulted.

2.1.1 Primary objective

To compare, in patients with cancer of the oesophagus and oesophago-gastric junction, the clinical and cost-effectiveness of minimally invasive laparoscopically-assisted (LAO) and open (OO) oesophagectomy procedures in terms of recovery, health-related quality of life, cost and survival.

2.1.2 Secondary objectives

To prospectively document the outcomes of totally minimally invasive oesophagectomy (TMIO) in an embedded study, with patients meeting the study eligibility criteria.

2.2 Trial design and configuration

The core of ROMIO is a multi-centre individually 1:1 randomised two-parallel group surgical trial.

Two centres are randomising 1:1:1 between three groups, the third being TMIO. The analysis of data from the TMIO arm will be conducted separately to the analysis of the core two-group trial, with a focus on fidelity, safety, and any signs of "promise" for the fully minimally invasive procedure increasing the rate of recovery of quality of life.

2.3 Trial centres

At the time of writing Bristol (+ Bath recruitment centre – MDT and surgery at Bristol), Plymouth, Leicester, Edinburgh, Preston, Nottingham, Salford (+ merger with South Manchester) and Southampton are open to recruitment.

2.4 Eligibility criteria
2.4.1 Inclusion criteria (abridged)

18 years of age or older

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

Referred for primary oesophagectomy by the MDT or oesophagectomy following re-staging after neodjuvant treatment

Confirmed MDT evidence of at least adenocarcinoma or at least squamous cell cancer of the oesophagus or oesophago-gastric junction

Fit for pre-operative anaesthesia and surgery, assessed by the MDT

Measurement that the tumour starts more than 5cm below the crico-pharyngeus

Measurement that the tumour involves less than 4cm of the gastric wall

The final pre-treatment tumour stage is between T1N0M0 and T4aN1M0.

2.4.2 Exclusion criteria (abridged)

Patients with high grade dysplasia

Tumour stage is T4b, or evidence of metastatic disease

Type 3 tumours of the oesophago-gastric junction that are scheduled for total gastrectomy

Patients with squamous cell cancer of the oesophagus who the MDT recommends or who individually elect to undergo definitive chemoradiotherapy

Evidence of previous complex thoracotomies or laparotomies that preclude a minimal access approach

Evidence of previous/concomitant malignancy that would interfere with this treatment protocol

Pregnancy

Patients participating in other trials that would interfere with the implementation of this protocol at a particular site.

2.5 Description of interventions

Oesophagectomy consists of a two-field lymphadenectomy: abdomen and thorax. All study patients in the core two-arm trial undergo the same surgery but are randomly allocated to one of two methods of access.

2.5.1 Open oesophagectomy (OO)
The following approaches are permitted: two-phase (right thoracotomy, laparotomy), three-phase (right thoracotomy, laparotomy, cervical incision) or left thoracoabdominal. Within these boundaries, the location and length of incisions are at each surgeon's discretion. Methods to close the incisions are also at the surgeon's discretion.

2.5.2 "Laparoscopically assisted" oesophagectomy (LAO, a.k.a. "hybrid")

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

As above, except that access to the abdominal cavity will be achieved with several small (~12mm) incisions (as many as needed) and surgery performed laparoscopically. A larger incision to create a feeding jejunostomy is allowed, but should be less than 8cm in length.

2.6   Randomisation procedures

Randomisation will be carried out after eligibility has been confirmed and consent given. Every effort should be made for surgery to be carried out within two weeks of randomisation; adherence to this will be monitored.

Randomisation will be performed by an authorised member of the local research team using a secure internet-based randomisation system ensuring allocation concealment and the avoidance of selection bias.

Allocation of patients to LAO or OO will be at random, will be conducted separately for each centre, and further stratified by whether the patient has undergone neoadjuvant treatment or not. Randomisation within blocks of varying size will prevent large imbalances in the number of patients in each treatment group, whilst maintaining allocation concealment.

2.7   Sample size and justification

The theoretical advantage of LAO compared to OO for patients is improved short-term recovery with the long-term survival benefit of surgery maintained. Consequently, the primary endpoint is the QLQ C30 Physical Function sub-scale measured at three and six-weeks post-surgery and three months post-randomisation.

For simplicity, and to indicate the minimum statistical power that will be achieved for the comparison of recovery, we consider just the six-week assessment of physical function. The planned analysis, based on the three-week, six-week, and three-month assessments of patient-reported physical function (primary outcome), and the baseline assessment as a covariate, is likely to have greater power than indicated here.

We are assuming that having adjusted analyses for centre, there will be no further need to accommodate clustering of outcomes by surgeon. In fact, as a team of surgeons is involved in each case (in decision-making, in-hospital care, and often in theatre), it would be difficult to do in practice.

A recent review of patient reported outcomes has indicated that the minimum clinically important difference on the QLQ-C30 Physical Function Scale is 0.4 standard deviations (1). Allowance for 5% of patients allocated to LAO undergoing OO, and 10% of patients in each group being found during surgery to have more extensive disease, can be achieved by reducing the effect size to be detected to 0.34 standard deviations. In this situation 182 patients in each group (364 patients in total) will allow a true treatment effect (LAO versus OO) of 0.4 standard deviations to be detected with 90% power at the 5% significance level, when up to 15% of patients are not able to follow their allocated procedure.

Further allowing for up to 10% missing outcome data, e.g. due to the patient being too sick, increases the target sample size to 364/0.9 = 406 patients in total. Hence our sample size target for the definitive ROMIO trial is 203 patients allocated to LAO and 203 patients allocated to OO.

We agreed with the NIHR on the 24th January 2019 to extend recruitment until September 2019 to allow 300 patients to be recruited to the main study. The remaining (approx. 120) patients will come from the external feasibility study cohort. This is treating the feasibility study as an internal pilot, this being possible as it had the same design as the main trial, and it continued to recruit whilst preparations were made for the main trial.

2.8   Blinding

In the first week post-surgery patients will be blinded using large adhesive dressings, positioned similarly on all trial patients regardless of the type of surgery (covering abdominal, thoracic, and cervical incisions). The first dressing should be applied by the surgical team in the operating theatre. The dressing will not be changed unless required; it will then be changed according to local practice. If dressings are changed, the patient will be asked to turn their head away from the wound sites to prevent them from observing the wounds. The nurse will clean the sites of all actual and potential incisions. Dressings will be removed as per local practice at one-week post-surgery.
The surgical team will not be blind to allocation.

2.9   Trial committees

ROMIO has a Study Steering Committee (SSC, Independent Chair: Craig Ramsay, Aberdeen University) and a Data Monitoring and Safety Committee (DMSC, Independent Chair: Judith Bliss, Institute of Cancer Research, London).

The study statistician will provide unblinded analyses, by group, of outcomes and safety measures to closed sessions of the DMSC; no other members of the study management group will see unblinded outcome data (all data collected post-randomisation, unless a case can be made that a measure is unrelated to clinical outcome) prior to the writing of planned publications.

Unblinded analyses by group, for the DMSC, should be restricted to those measures required to monitor patient safety in the trial. Analyses of the clinical outcomes by group, including the primary outcome, should not be conducted routinely unless these are necessary for monitoring patient safety.

Once the study statistician has seen unblinded data, they should not be involved in Study Management Group (SMG) discussions of changes to study conduct.

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

2.10 Outcome measures

2.10.1 Primary outcome

The primary outcome will be the three assessments of physical function (a subscale of the EORTC QLQ-C30) assessed at three and six-weeks post-surgery and three months after randomisation.

2.10.2 Secondary outcomes

Secondary outcomes will assess the efficacy of the two approaches (morbidity and safety) and establish oncological markers of quality assurance of surgery which are surrogate markers of long-term survival (detailed histopathology and quality assurance of the radicality of surgery). Secondary outcomes will include:

1. All-cause short and long-term complications

2. Impact of complications up to hospital discharge will be categorised using the Clavien-Dindo System(2).

3. Spirometry measures of forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC)

4. Success of blinding during the first six days post-surgery, using the Bang Blinding Index(3).

5. Generic and disease specific HRQL measures EORTC QLQ-C30(4), and QLQ-OES18(5), multidimensional fatigue inventory (MFI-20)(6), EuroQoL EQ-5D-5L(7) from which we will calculate quality adjusted life years (QALYs).

6. Quality assurance of surgery with histopathological and surgical measures

7. Overall and recurrence-free survival to two years

8. Length of hospital stay, defined as length of primary hospital stage plus readmission within 30 days (and length of primary hospital stay plus length of hospital stay if discharged to community hospital)

9. Further measures of NHS resource use and costs.

The assessment points are pre-surgery, post-surgery at three days, six days, three weeks, and six weeks, and post-randomisation at three, six, nine, twelve, eighteen and 24 months.

2.11   Interim analysis

The study team have not planned any interim analysis of the outcomes data.

An unblinded analysis of complications, collected during the feasibility study, was presented to a closed session of the DMC at the committee's request (25th January 2018).

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

3.   GENERAL ANALYSIS CONSIDERATIONS
3.1   Analysis populations

Each of the following will include participants of the feasibility and main trials, according to the availability of data.

The primary effectiveness analysis will include data from the **"full analysis set"**; all randomised participants **in the groups to which they were allocated** (i.e. according to the intention-to-treat principle) who provide at least one of the three post-treatment assessments of physical function contributing to the primary outcome.

The **"safety set"** will include all randomised participants for whom there was an attempt to complete a study procedure (OO or LAO). This does not include patients for whom curative surgery did not proceed (e.g. due to the discovery of more advanced disease). The safety set does include participants who underwent a study procedure which was not that allocated, and participants who needed to convert from LAO to OO due to clinical reasons. This is the analysis population that will be used in the reporting of complications and adverse events. The groups to be compared will be defined by the study procedure **first initiated**, e.g. participants will be retained in the LAO group, even if they subsequently needed to convert to OO.

The **"per protocol set"** will include all randomised participants, **in their allocated groups**, for whom the allocated surgery was completed. Analyses of blinding, and surgical quality will use data from these participants.

Within the above three sets, about 120 feasibility phase patients, allocated to OO or LAO and with the necessary data, will be distinguished from those patients allocated to OO and LAO in the main study. In general, estimates of effectiveness will be stratified by feasibility and main study, and presented with a third pooled estimate using data from the two study phases.

3.2   Derived variables

Results based on derived variables will be sense-checked by the chief investigator. In addition, the code will be checked by the senior statistician.

3.3   Procedures for missing data

The primary effectiveness analysis will be based upon those randomised participants who provide at least one of the three post-treatment assessments of physical function contributing to the primary outcome.

The five QLQ-C30 items contributing to the physical function scale are as follows:

1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?
2. Do you have any trouble taking a long walk?
3. Do you have any trouble taking a short walk outside of the house?

4. Do you need to stay in bed or a chair during the day?
5. Do you need help with eating, dressing, washing yourself or using the toilet?

A patient will be considered to have completed the physical function scale and to have a score of zero (lowest level of physical functioning), if their reason for not having a physical function score is recorded as due to being in intensive care or having died prior to the questionnaire being due. This is a reasonable assumption about responses to the above five questions in those circumstances and is to prevent these values being implicitly imputed by the statistical model upon which the analysis is based. This approach will only be used for the physical function and role function scales, where the assumption of a zero value can be justified; the number of values imputed in this way will be reported by treatment group. The two items contributing to role function are:

6. Were you limited in doing either your work or other daily activities?
7. Were you limited in pursuing your hobbies or other leisure time activities?

Participants missing their baseline assessment of a patient reported outcome will be included in the analysis by (i) imputing the overall mean for the corresponding observed baseline measurements, (ii) including an indicator variable as a covariate with values 0 if the individual has provided a baseline measure, and value 1 if not. This approach has been shown to give unbiased estimates in a randomised trial, and to avoid the loss of power that a complete case analysis will cause(8).

When single items are missing from any of the patient reported outcomes, these will be addressed using the accepted guidelines where available.

3.4   Study centre effects

Variation between centres will be accommodated in each analysis by including dummy variables, distinguishing each centre (Bristol and Bath, plus Salford and South Manchester will be combined) in the regression model. There is no plan to investigate variation in the treatment effect by study centre.

3.5   Visit windows

Late completion of assessments may prove unavoidable in a small number of cases, but for the data to be accepted the assessments must be completed:

Pre-surgery: before the day of surgery

Post-surgery:

3 days: no earlier than day 2 and no later than day 4

6 days: no earlier than day 5 and no later than day 9

| 3 weeks: | no earlier than day 10 and no later than day 34 |
| 6 weeks: | no earlier than day 35 and no later than day 66 |

Post-randomisation:

3 months:	no earlier than day 67 post-surgery, and no later than day 111
6 months:	no earlier than 5 months and no later than 7 months
9 months:	no earlier than 8 months and no later than 10 months
12 months:	no earlier than 11 months and no later than 14 months
18 months:	no earlier than 15 months and no later than 20 months
24 months:	no earlier than 21 months and no later than 30 months

Wide windows for completion have been allowed, so that all data obtained from the longer-term follow-up can be used. However, efforts will be made to encourage completion within 10 days of each assessment point. Variation in the timing of completion will be accommodated by the statistical analyses. As the data from the feasibility study are now contributing to the primary analysis, these time windows will be applied to feasibility and main studies – this is a slight change to the windows in the feasibility study protocol.

4. DESCRIPTION OF PARTICIPANT CHARACTERISTICS

4.1 Disposition

The flow of patients through the trial will be summarised in a CONSORT diagram that will include the eligibility, reasons for exclusion, numbers randomised to the two treatment groups, those undergoing their allocated surgery, losses to follow up and the numbers analysed (see Figure 1 in Section 8 of this plan).

4.2 Baseline characteristics

Summary statistics for patient characteristics, as determined at baseline, will be presented separately for those subsequently allocated to LAO and OO. The patient characteristics presented will allow the ROMIO patients to be compared to cohorts in other studies of oesophagectomy (see Table 1 in Section 8 of this plan).

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

5. ASSESSMENT OF STUDY QUALITY

5.1 Eligibility checks

The numbers of patients and reasons for exclusion will be described; this information will be presented in the CONSORT flow chart.

5.2 Participating surgeons

The number of surgeons participating at each centre, and confirmation that they met the entry criteria for participation in ROMIO, will be presented in the text of the main reports.

5.3 Study completion

All patients will be followed up for 24 months post-randomisation, for vital status, and through completion of PROs. If possible within the study follow-up period, patients can be followed up for 36 months.

5.4 Compliance

The number of patients who, post-randomisation, opt for surgery other than that allocated, will be reported in the CONSORT flow chart.

5.5 Fidelity to allocated surgery

Surgeons are completing check-lists for, and photographing key aspects of each procedure, to assess whether they are abiding by the listed mandatory and prohibited actions for OO, and for LAO. The number of procedures for which fidelity was maintained will be reported by study arm for patients in the per protocol set, and the nature of the deviation presented by study group for other patients.

Conversions from allocated surgery to another approach, and cases where the allocated surgery could not be completed, will be reported in the CONSORT flow chart (these patients are not part of the per protocol set) and reasons noted in the text.

5.6 Changes made to the statistical analysis plan

Subsequent versions of the statistical analysis plan will include a table of changes made since the previous version. When the plan cannot be followed in the primary effectiveness analysis, infringements of the plan will be detailed with reasons in the report of primary results.

6. ANALYSIS OF EFFECTIVENESS

6.1 Patients found to be ineligible post-randomisation

For patients found to be ineligible post-randomisation, the reason for ineligibility will be noted, and the patient invited to continue in follow-up. Where the required data are provided, the patient will be included in the full analysis set for the intention to treat analysis of effectiveness.

6.2 Summary of primary and secondary outcomes

For patients in the full analysis set, the number of patients providing data, mean, and standard deviation will be presented, by allocated group, for the time points used in the analysis of the primary outcome: baseline, three and six-weeks post-surgery, and three months post-randomisation. The primary outcome is based on the physical function scale of the QLQ-C30, a continuous measure with high scores indicating better physical function.

With regards to secondary outcomes, the same statistics will be presented for other scales derived from patient reported outcomes:

> For function scales derived from the QLQ C-30 (role, emotional, cognitive, social), and global health status, high scores indicate high functioning / good quality of life.

> For symptom scales derived from the QLQ C-30 (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea, financial difficulties), high scores indicate greater symptom burden.

> For the function scale derived from the QLQ-OES18 (dysphagia), a high score indicates good functioning.

> For the symptom scales and items derived from the QLQ-OES18 (eating, reflux, pain, trouble swallowing saliva, choked when swallowing, dry mouth, trouble with taste, trouble with coughing, trouble talking), a high score indicates a high symptom burden.

> For the five scales derived from the MFI-20 (general fatigue, physical fatigue, reduced activity, reduced motivation, mental fatigue) a high score indicates greater fatigue.

For the full analysis set, the occurrence of deaths, and of disease recurrence over a two-year follow-up period will each be presented for the LAO and OO groups as a Kaplan-Meier survival curves. Point estimates from these survival curves will be presented at 30 days and 90 days to allow comparison with National Audit Data.

For the safety set, the occurrence for each patient of one or more complications in each of the following key categories, both up to hospital discharge, and separately for the remainder of the 24-month follow-up, will be tabulated for the LAO and OO groups: oesophagoenteric leak and conduit necrosis/failure (patients experiencing each and both of these will be reported), chyle leak, pneumonia, gastrointestinal bleeding requiring intervention or transfusion. The number of patients requiring readmission to ITU/HDU will also be presented.

For the per protocol set, summary statistics for the histopathological and surgical measures of quality will be presented for each of the LAO and OO groups, including length of oesophagus, total count of malignant 'positive' nodes, total count of all nodes, carcinoma positive circumferential resection margins, carcinoma positive proximal and distal resection margins, and pT stage.

6.3   Primary analysis

The null hypothesis being addressed by the primary analysis is that, in truth, physical functioning in the three months following randomisation, measured with the physical function scale of the QLQ-C30, is identical whether access for a patient's oesophagectomy was achieved with LAO or OO.

Using the full analysis set, the following analysis will be conducted separately for feasibility and full trial patients:

The relative effectiveness of the two approaches will be quantified as a difference in mean response on the physical function scale of the QLQ-C30 (LAO mean – OO mean), amongst patients in the full analysis set, using patients' available measures from the three-week and six-week post-surgery, and the three-month post-randomisation assessments. The difference in mean response will be presented with its 95% confidence interval, and p-value.

The difference in means will be estimated in a mixed effects linear regression model with patient response at three-week (y_{i1}) and six-week (y_{i2}) post-surgery, and three-month (y_{i3}) post-randomisation as the outcome variables (i.e. between one and three outcome measurements per patient), and covariates: treatment allocation (x_{1i}=1: LAO; x_{1i}=0: OO), baseline physical function (x_{2i}), dummy variables distinguishing centres (x_{3ki}=0,1; k=1,...K indexing the centres), two dummy variables distinguishing assessment points (x_{4ti}=0,1; t=1,2,3), and a dummy variable distinguishing patients who underwent neoadjuvant treatment or not (x_{5i}=1: underwent neoadjuvant treatment; x_{5i}=0: no neoadjuvant treatment). Finally a dummy variable distinguishing those participant without a baseline assessment of outcome (x_{6i}=1: no baseline assessment; x_{5i}=0: baseline assessment available).A normally distributed random effect will accommodate the correlation between each patient's responses: z_i~N(0,σ_z). A normal distribution is assumed for the residual errors: e_i~N(0,σ_e). The coefficient for the treatment allocation covariate (β_1) is the intention to treat estimate of treatment effectiveness, comparing LAO to OO. In statistical notation:

$$y_{it} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sum_{k=1}^{K} \beta_{3k} x_{3ki} + \sum_{t=1}^{3} \beta_{4t} x_{4ti} + \beta_5 x_{5i} + \beta_6 x_{6i} + z_i + e_i$$

The estimates from the feasibility and full trial patients will both be presented, with a third pooled estimate computed as a mean of the feasibility and full trial estimates, weighted by the inverse of the variance of the estimate (equivalent to a fixed-effects meta-analysis). **This pooled estimate is the primary estimate of relative treatment effectiveness.**

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

6.4 Secondary analyses

6.4.1 Patient reported outcome questionnaires

Using the full analysis set, the primary analysis plan will be adapted to the analysis of each of the other scales from patient reported outcomes. For these secondary analyses, it may be necessary to present the pooled estimate (combining feasibility and full trial estimates) in the paper, with the feasibility and full trial specific estimates presented in supplementary material. Where the feasibility and full trial estimates are noticeably different to the extent of supporting different conclusions, this will be highlighted.

6.4.2 Quality assurance of surgery

Using the per protocol set, the null hypothesis of true equality of quality of surgery being achieved with LAO and OO will be tested for each measure in turn. As the per protocol set is being employed here, conclusions will need to be cautious, and only simple tests will be used: e.g. t-tests for normally distributed continuous measures, Mann-Whitney U tests for markedly non-normal continuous measures, and chi-square tests for binary measures.

6.4.3 pStage based on resection

Staging based on the resection will be categorised according to the American Joint Committee on Cancer 8th Edition.(9)

6.4.4 Surgical complications during hospital stay

Using the safety set, the null hypothesis of true equality in the proportion of patients experiencing one or more episodes of each of the key complications (oesophagoenteric leak, conduit failure, chyle leak, pneumonia, GI bleeding requiring intervention) will be tested using a chi-square test.

For each patient the most severe impact of any complications during their initial hospital stay will be categorised using the Clavien-Dindo system(2), patients in the LAO and OO groups being compared using ordered logistic regression.

6.4.5 Post-operative pain

Using the full analysis set, the number responding, median, first and third quartile pain scores will be presented for LAO and OO groups at each of three days and six days. Medians and first and third quartiles will be presented. The null hypothesis of no true difference in pain at 6 days, comparing LAO and OO groups will be tested using the Mann-Whitney U test. Where patients have responded with a score between 0 and 10 rather than completing the visual analogue scale, the score will be converted by assuming that each point score = 10mm along the visual analogue scale.

6.4.6 Post-operative lung function

Using the full analysis set, for each of the three- and six-day assessments separately, FEV$_1$ and FVC as a percentage of the baseline measurement will be compared between the LAO and OO groups. Medians and first and third quartiles will be presented. The evidence against the null hypothesis of no difference in lung function between OO and LAO groups will be quantified using the Mann-Whitney U test.

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

6.4.7 Length of hospital stay

Using the full analysis set, total and post-surgical inpatient stay will be compared between LAO and OO groups. Medians and first and third quartiles will be presented. The evidence against the null hypothesis of no difference in hospital stay between OO and LAO groups will be quantified using the Mann-Whitney U test.

6.4.8 Overall and recurrence-free survival to two years.

A Kaplan-Meier plot will present survival over time in the OO and LAO groups. Kaplan-Meier estimates of survival, with 95% confidence intervals will be presented for survival in each of the OO and LAO groups at the following time points: 30 days, 90 days, 24 months.

Using the full analysis set, the hazard ratio comparing those allocated to LAO to those allocated to OO in terms of all-cause mortality, and recurrence-free survival will be estimated using proportional hazards models with time to event as the outcome measure and covariates: allocated surgery, dummy variables distinguishing the centres, and a dummy variable distinguishing those patients who underwent neoadjuvant treatment, and those who did not. Marked evidence against the proportional hazards assumption will be investigated in a test based on Schoenfeld residuals(10). Estimated hazard ratios will be presented with 95% confidence intervals.

This analysis is aimed at ruling out a large difference in survival between LAO and OO, as the benefit of LAO is expected to be limited to faster recovery in the initial months post-surgery. Our equivalence bound is a 7% difference in mortality risk, such that the 95% confidence interval for the risk difference at 24 months post-randomisation is entirely within the range -7% to +7%.

6.4.9 Resource use measures & the EQ-5D-5L

Resource use (e.g. procedure costs, length of stay, readmission etc) QALYs (derived from EQ-5D-5L and mortality data) and cost-effectiveness analyses will be detailed in a separate health economics plan.

6.5 Subgroup analysis

Sub-group analyses will investigate whether the relative effects of OO and LAO differ according to:

- Whether a patient underwent neoadjuvant chemotherapy/chemoradiotherapy versus no neoadjuvant treatment prior to surgery.
- POSSUM physiological score assessed at recruitment.
- BMI assessed at recruitment.

Separate treatment effect estimates and 95% confidence intervals will be presented for those undergoing neoadjuvant treatment and those who do not, and for those above or at the median and those below the median POSSUM score and BMI. The evidence for a modification of the true treatment by each of the above three measures in turn will be quantified by adding an interaction term to the above primary analysis model. For neoadjuvant treatment the interaction term will be a binary covariate distinguishing those

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

participants who underwent any neoadjuvant treatment and were allocated to the LAO group. For POSSUM score and BMI, the interaction term will be equal to the POSSUM score or BMI if the participant is in the LAO group, zero otherwise.

6.6 Sensitivity analysis

If, at one or more of the three time points contributing to the primary outcome, more than 20% of data is missing, the recorded reasons for missing data will be assessed, and data at the three time points imputed in accordance with those reasons. This imputed dataset will be the basis of a sensitivity analysis.

As participants whose surgery could not be completed are likely to be informed of this, the analysis of pain at six days post-surgery will be repeated using the per protocol set.

The primary outcome requires that the average time from randomisation to surgery is the same for the study groups being compared. We will conduct a sensitivity analysis, additionally adjusted for each patient's time from randomisation to surgery.

6.7 Exploratory/other analysis

For each scale from the patient reported outcomes, the mean response for LAO and OO groups at each assessment point up to 24 months will be presented graphically, with an indication of the number of patients in each group contributing to the estimates. Differences in quality of life are not expected to persist following the post-surgical recovery period. Any observed differences subsequent to three months post-randomisation will be noted in this exploratory analysis.

The Bang Blinding Index will be presented for both LAO and OO groups separately to assess the success of patient blinding to allocation in the first week post-surgery.

7.   PUBLICATION PLAN

Once all allocated patients have completed their three-month assessment (late 2019) a data extract will be prepared for analysis of the main peri- and post-surgical measures, including the primary outcome. The tables and figures for this paper are presented in the next section (Tables 1, 2, 3, 4, Supplementary Table 1, Figure 1).

Once all allocated patients have completed their 24-month assessment (autumn 2021) a further data extract will be prepared for analysis of the PROs, complications and survival over that longer period (Table A, Figures A, B, C, D, E, Supplementary Tables as detailed in the next section).

8. FIGURES AND TABLES FOR JOURNAL PUBLICATION OF PRIMARY ANALYSIS

The following are the intended figures and tables for the main journal publications of the ROMIO results. Numbered tables and figures will be presented in the initial paper, those indexed by letters will appear in the subsequent paper.

Table 1. Baseline Characteristics by allocated surgery

	OO (n=)	LAO (n=)
Male; n (%)		
Mean age (SD), n		
Mean BMI (SD), n		
Mean POSSUM physiological score (SD), n		
WHO performance status score, n (%)		
0		
1		
2		
Tumour histologic findings; n (%)		
Adenocarcinoma		
Squamous cell carcinoma		
Adenosquamous		
Other		
Location of tumour in oesophagus; n (%)		
Upper third		
Middle third		
Lower third / Siewert I		
Junctional tumour, Siewert II		
Junctional tumour, Siewert III		
Clinical tumour classification; n (%)		
cT1		
cT2		
cT3		
cT4a		
Clinical node classification; n (%)		
cN0		
cN1		
cN2		
Neoadjuvant treatment; n (%)		
Chemotherapy		
Chemoradiotherapy		
None		

Table 2. Post-surgical recovery of physical function by allocated surgery (primary analysis)

	OO (n=)	LAO (n=)
PRIMARY ANALYSIS – pooled estimate		
Pre-randomisation mean (SD) n		
Three-week post-surgery mean (SD) n		
Six-week post-surgery mean (SD) n		
Three-month post-randomisation mean (SD) n		
Treatment effect (95% confidence interval) p-value		
SENSITIVITY ANALYSES		
Adjusted for time from randomization to surgery		
Treatment effect (95% confidence interval)		
Missing outcome data imputed		
Treatment effect (95% confidence interval)		

Table 3. Complications in those participants undergoing allocated surgery. Counts are of patients, who may have experienced more than one complication of a given type

	OO (n=)	LAO (n=)	p-value
Clavien-Dindo Classification of Complications, n (%)			
Normal recovery			
Grade I/II			
Grade IIIa / Grade IIIb			
Grade IVa / Grade IVb			
Grade V (death of patient)			
Key post-operative complications within 31 days of surgery, n (%)			
Oesophagoenteric leak from anastomosis, staple line, or localised conduit necrosis			
Conduit necrosis/failure			
Chyle leak			
Pneumonia / chest infection			
Delayed gastric emptying			
GI bleeding requiring intervention or transfusion			
Acute abdominal wall dehiscence/hernia			
Acute diaphragmatic hernia			
Key post-operative complications between 31 and 90 days of surgery, n (%)			
Oesophagoenteric leak from anastomosis, staple line, or localised conduit necrosis			
Conduit necrosis/failure			
Chyle leak			
Pneumonia / chest infection			
Delayed gastric emptying			
GI bleeding requiring intervention or transfusion			
Acute abdominal wall dehiscence/hernia			
Further intervention, n (%)			
Reoperation during index hospitalisation			
Return to ITU/HDU during index hospitalisation			
Within three months of randomisation			

Table 4. Peri- and post-surgical measures

	OO (n=)	LAO (n=)	p-value
30-day post-operative mortality, n (%)			
90-day post-operative mortality, n (%)			
Median total hospital stay in days (Q1, Q3)[1]			
Median post-surgical hospital stay in days (Q1, Q3)[1]			
Median total lymph nodes retrieved (Q1, Q3)[1]			
Median lymph nodes with tumour (Q1, Q3)[1]			
Resection margins, n (%)[2]			
R0			
R1			
R2			
pStage, n (%)[2]			
0			
I or II			
III			
IV			
No residual tumour or lymph-node metastasis			
Median FEV$_1$, as a percentage of pre-surgery (Q1 Q3,), n			
Three days post-surgery[1]			
Six days post-surgery[1]			
Median FVC, as a percentage of pre-surgery (Q1 Q3,), n			
Three days post-surgery[1]			
Six days post-surgery[1]			
Median pain (Q1, Q3), n; 0 is no pain, 100 is worst possible pain.			
Three days post-surgery[1]			
Six days post-surgery[1]			

1. p-value based on Mann-Whitney U statistic. Note that Q1 and Q3 are the first and third quartiles
2. p-value based on ordered logistic regression

Supplementary Table 1. Post-surgery treatment complications in those undergoing their allocated surgery [specific complications will be detailed if they occur in one or more patients – those listed here are examples only]

	OO (n=)	LAO (n=)
Pulmonary complications, n (%)		
Respiratory failure requiring intubation		
Acute respiratory distress syndrome		
Cardiac complications, n (%)		
Dysrhythmia atrial requiring treatment		
Dysrhythmia ventricular requiring treatment		
Urological complications, n (%)		
Acute renal insufficiency		
Gastrointestinal complications, n (%)		
Anastomotic stricture requiring endoscopic intervention		
Thromboembolic complications, n (%)		
Pulmonary embolism		
Neurological / psychiatric complications, n (%)		
Other neurological injury		
Post-operative infections, n (%)		
Wound infection requiring opening wound or antibiotics		
Post-operative wound / diaphragm complications, n (%)		
Acute abdominal wall dehiscence / hernia		
Other post-operative complications, n (%)		
Multiple organ dysfunction syndrome		

Supplementary Table 2. Cohort stratified and pre-specified subgroup analyses

	OO (n=)	LAO (n=)
Main study cohort		
Pre-randomisation mean (SD) n		
Three-week post-surgery mean (SD) n		
Six-week post-surgery mean (SD) n		
Three-month post-randomisation mean (SD) n		
Treatment effect (95% confidence interval)		
Feasibility study cohort		
Pre-randomisation mean (SD) n		
Three-week post-surgery mean (SD) n		
Six-week post-surgery mean (SD) n		
Three-month post-randomisation mean (SD) n		
Treatment effect (95% confidence interval)		
Neoadjuvant treatment		
Treatment effect (95% confidence interval)		
No neoadjuvant treatment		
Treatment effect (95% confidence interval)		
	Interaction p-value:	
BMI at or above the median		
Treatment effect (95% confidence interval)		
BMI below the median		
Treatment effect (95% confidence interval)		
	Interaction p-value:	
POSSUM Physiology at or above the median		
Treatment effect (95% confidence interval)		
POSSUM Physiology below the median		
Treatment effect (95% confidence interval)		
	Interaction p-value:	

Figure 1: CONSORT Flow chart

```
┌─────────────────────────────────┐     ┌──────────────────────────────────┐
│ Referred by Multi-disciplinary  │     │ Excluded N =                     │
│ team for primary oesophagectomy │     │ Declined randomisation N =       │
│ or neoadjuvant treatment and    │     │ Not eligible N =                 │
│ surgery                         │     │ General health worsened N =      │
│            N =                  │     │ Participating in competing trial │
│                                 │     │   N =                            │
│                                 │────▶│ Lacked capacity to consent N =   │
│                                 │     │ Non-recruiting Trust N =         │
│                                 │     │ Administration failure N =       │
│                                 │     │ Not recorded N =                 │
└─────────────────────────────────┘     └──────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│           Randomised            │
│             N =                 │
└─────────────────────────────────┘
              │
      ┌───────┴────────┐
      ▼                ▼
┌──────────────────┐  ┌──────────────────┐
│ Allocated open   │  │ Hybrid surgery   │
│ surgery N =      │  │ N =              │
│ Underwent open   │  │ Underwent open   │
│  surgery n=      │  │  surgery n=      │
│ Underwent hybrid │  │ Underwent hybrid │
│  surgery n=      │  │  surgery n=      │
│ Other n=         │  │ Other n=         │
└──────────────────┘  └──────────────────┘
      │                │
      ▼                ▼
┌───────────────────────────────────────────────────────┐
│ QLQ C-30 Physical function completed at least one of   │
│ six days, six weeks and three months assessment points │
└───────────────────────────────────────────────────────┘
      │                │
      ▼                ▼
┌──────────────────┐  ┌──────────────────┐
│ Included in      │  │ Included in      │
│ primary analysis │  │ primary analysis │
│ N=               │  │ N=               │
└──────────────────┘  └──────────────────┘
```

University of BRISTOL

CTEU Bristol

Bristol Randomised Trials Collaboration

Table A. Survival and recurrence-free survival

	OO (n=)	LAO (n=)
Survival over 24 months post-randomisaton		
Person-years of follow-up per group over 24 months		
Number of events over 24 months		
Survival at 24 months (95% confidence interval)		
Survival over full follow-up period		
Total person-years of follow-up per group		
Number of events		
Hazard ratio (95% confidence interval) p-value		
Recurrence-free survival over 24 months post-randomisation		
Person-years of follow-up per group over 24 months		
Number of events over 24 months		
Survival at 24 months (95% confidence interval)		
Recurrence-free survival over full follow-up period		
Total person-years of follow-up per group		
Number of events		
Hazard ratio (95% confidence interval) p-value		

Figure A: Mean QLQ-C30 Function sub-scale scores over two years of follow-up in patients allocated to LAO and OO. High scores indicate good function

Figure B: Mean QLQ-C30 Symptom sub-scale scores over two years of follow-up in patients allocated to LAO and OO. High scores indicate severe symptoms

Figure C: Mean QLQ-OES18 Function and Symptom sub-scale scores over two years of follow-up in patients allocated to LAO and OO. A high score indicates good swallowing function, and high symptom burden on the other measures

Figure D: Mean MFI-20 sub-scale scores over two years of follow-up in patients allocated to LAO and OO. High scores indicate greater fatigue

Figure E: Kaplan-Meier plot of (a) survival (b) recurrence-free survival in patients allocated to LAO and OO

Supplementary Tables: will give the numeric estimates of the means (standard deviations) for the PRO measures. An example table is given here to show the structure.

	OO (n=)	LAO (n=)
	Mean (SD) n	Mean (SD) n
QLQ-C30 Physical Function		
Pre-randomisation		
6-day post-surgery		
3-week post-surgery		
6-week post-surgery		
3-month post-randomisation		
6-month post-randomisation		
12-month post-randomisation		
18-month post-randomisation		
24-month post-randomisation		
QLQ-C30 Role Function		
Pre-randomisation		
6-day post-surgery		
3-week post-surgery		
6-week post-surgery		
3-month post-randomisation		
6-month post-randomisation		
12-month post-randomisation		
18-month post-randomisation		
24-month post-randomisation		

Statistical Analysis Plan
ROMIO

University of
BRISTOL

CTEU
Bristol

Bristol
Randomised Trials
Collaboration

Supplementary Table X. Complications between three months and two years post-randomisation in those undergoing their allocated surgery [specific complications will be detailed if they occur in one or more patients – those listed here are examples only]

	OO (n=)	LAO (n=)
Other pulmonary complications, n (%)		
Respiratory failure requiring intubation		
Acute respiratory distress syndrome		
Cardiac complications, n (%)		
Dysrhythmia atrial requiring treatment		
Dysrhythmia ventricular requiring treatment		
Urological complications, n (%)		
Acute renal insufficiency		
Other gastrointestinal complications, n (%)		
Anastomotic stricture requiring endoscopic intervention		
Thromboembolic complications, n (%)		
Pulmonary embolism		
Neurological / psychiatric complications, n (%)		
Other neurological injury		
Post-operative infections, n (%)		
Wound infection requiring opening wound or antibiotics		
Post-operative wound / diaphragm complications, n (%)		
Acute abdominal wall dehiscence / hernia		
Other post-operative complications, n (%)		
Multiple organ dysfunction syndrome		

9. DESCRIPTIONS OF CLINICAL MEASURES

9.1 WHO Performance Status

0:    Fully active, able to carry on all pre-disease performance without restriction

1:    Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work

2:    Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours

9.2 POSSUM Physiological Score

The POSSUM Physiological score(11) is derived from the following physiological measures pre-surgery, with higher scores indicating poorer health:

Age in years
Cardiac signs
Respiratory history
Chest radiograph
Systolic blood pressure
Pulse
Glasgow coma score
Haemoglobin
White cell count
Urea
Sodium
Potassium
Electrocardiogram

9.3 Siewart Classification

The Siewert-Stein classification is a system of anatomical classification used for adenocarcinomas of the esophagogastric junction.(12)

Type 1:   Adenocarcinoma of the distal part of the esophagus. The tumour center is located 1-5 cm above the gastric cardia.

Type 2:   Adenocarcinoma of the real cardia. The tumour center is located 1cm above or 2cm below the gastric cardia. Considered to be true gastroesophageal junction.

Type 3:   Adenocarcinoma of the subcardial stomach. The tumor center is located 2–5 cm below the gastric cardia

9.4 TNM classification, 7th edition

Primary Tumour
TX: Primary tumour cannot be assessed
T0 :   No evidence of primary tumour

Tis: Carcinoma in situ / high grade dysplasia
T1a: Tumour invades the lamina propria or muscularis mucosae
T1b: Tumour invades submucosa
T2 :   Tumour invades muscularis propria
T3 :   Tumour invades adventitia
T4a :   Tumour invades pleura, pericardium or diaphragm
T4b: Tumour invades other adjacent structures e.g. aorta, vertebral body or trachea

Regional Lymph Nodes
NX: Regional lymph nodes cannot be assessed
N0:   No regional lymph node metastasis
N1:   Metastasis in 1-2 regional lymph nodes
N2:   Metastasis in 3-6 regional lymph nodes
N3:   Metastasis in 7 or more regional lymph nodes

Metastasis
M0:   No distant metastasis
M1:   Distant metastasis

9.5   Clavien-Dindo Classification of Complications

Using the Clavien-Dindo system(2), each patient will be graded according to the most serious complication experienced:

0:    Normal recovery with no complications
I: Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic or radiological intervention
II: Requiring pharmacological treatment with drugs other than such allowed for Grade I complications
IIIa: Requiring surgical, endoscopic or radiological intervention. Intervention not under general anaesthesia.
IIIb: Requiring surgical, endoscopic or radiological intervention. Intervention under general anaesthesia.
IVa: Life-threatening complication (including CNS complications) requiring IC/ITU management. Single organ dysfunction.
IVb: Life-threatening complication (including CNS complications) requiring IC/ITU management. Multi organ dysfunction.
V: Death

9.6   R classification

Classification of resection margins. (9)
R0:   No residual tumour at resection margin
R1:   Microscopic residual tumour at or within 1mm of resection margin
R2:   Macroscopic residual tumour at resection margin

9.7 p-stage

A summary score for the excised tumour. (9) The following summary applies to both adenocarcinoma and squamous cell carcinoma, with the staging being the highest that applies.

IV: T4a + N2, any T4b, any N3, any M1
III: T4a + N0-1, T1-3 + N2, T2-3 + N1
I or II: All other T1a-3
0: Tis

10. REFERENCES

1.   Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol. 2011;29(1):89-96.

2.   Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009;250(2):187-96.

3.   Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. Control Clin Trials. 2004;25(2):143-56.

4.   Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85(5):365-76.

5.   Blazeby JM, Conroy T, Hammerlid E, Fayers P, Sezer O, Koller M, et al. Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. Eur J Cancer. 2003;39(10):1384-94.

6.   Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res. 1995;39(3):315-25.

7.   Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727-36.

8.   Groenwold RH, White IR, Donders AR, Carpenter JR, Altman DG, Moons KG. Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. CMAJ. 2012;184(11):1265-9.

9.   Rice TW, Patil DT, Blackstone EH. 8th edition AJCC/UICC staging of cancers of the esophagus and esophagogastric junction: application to clinical practice. Ann Cardiothorac Surg. 2017;6(2):119-30.

10.   Grambsch PM, Therneau TM. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. Biometrika. 1994;81(3):515-26.

11.   Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg. 1991;78(3):355-60.

12.   Siewert JR, Stein HJ. Classification of adenocarcinoma of the oesophagogastric junction. Br J Surg. 1998;85(11):1457-9.