Oral versus intravenous antibiotic treatment for bone and joint infections requiring prolonged antibiotic treatment: Multi-centre study

Statistical Analysis Plan

Version 2.0 - 03/12/2016

Based on version 2.0 - 01/05/2015 of protocol

	Name	Title/Role	Signature	Date
Author, approver	Ines Rombach	Trial Statistician		
Reviewer, approver	Sarah Walker	Senior statistician		
Approver	Matthew Scarborough	Chief Investigator		

RCS Surgical Intervention Trials Unit

CONTENTS

1. Introduction

This document details the proposed presentation and analysis for *the HTA-funded Multicentre Randomised Controlled Trial of Oral versus Intravenous Antibiotic Treatment for bone and joint infections requiring prolonged treatment (OVIVA)*. Any primary reporting of the OVIVA study should follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (for example, to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (for example, data transformation prior to analysis), but they are intended to establish the rules that will be followed, as closely as possible, when analysing and reporting the trial.

The analysis strategy will be available on request when the principal papers are submitted for publication in a journal. Suggestions for subsequent analyses by journal editors or referees, will be considered carefully, and carried out as far as possible in line with the principles of this analysis strategy; if reported, the source of the suggestion will be acknowledged.

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis should be carried out by an identified, appropriately qualified and experienced statistician, who should ensure the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

1.1 *Key personnel*

Trial statistician(s):

Ines Rombach (Surgical Intervention Trials Unit)

Chief Investigator:

Matthew Scarborough (Oxford University Hospitals)

Trial Manager:

Rhea Zambellas (Surgical Intervention Trials Unit)

Trial Physician:

Ho Kwong Li (Oxford University Hospitals)

DSMC Members:

Neil French, DMC chair, Professor of Infectious Disease, Liverpool University

Colette Smith, Lecturer in Biostatistics, UCL

Martin Llewelyn, Reader in Infectious Diseases and Therapeutics, Brighton and Sussex University

TSC Members:

Dr Graham Cooke, Consultant Infectious Disease and Senior Lecturer

Dr John Paul, Lead Public Health microbiologist, South East Region

Mr Fraser Old, retired

2. Changes from previous version of SAP

This is the second version of the statistical analysis plan, based on protocol version 2.0, 05th May 2015. Details on changes from previous versions are provided in section 13.

3. Background Information

3.1 *Objectives*

Primary Aim

To determine whether oral antibiotics are non-inferior to intravenous antibiotics for serious bone and joint infection judged by the percentage of patients experiencing definitive treatment failure during 1 year of follow up.

Secondary Aims

To compare the following endpoints according to treatment allocation;

1) SAEs, including death (i.e. all cause) according to treatment allocation.
2) line complications (i.e. infection, thrombosis or other events requiring early removal or replacement of the line).
3) *Clostridium difficile* associated diarrhoea
4) "probable" and "possible" treatment failure as composites with definitive treatment failure (see endpoint definitions and analysis section for details).
5) early termination of the planned 6 week period of oral or IV antibiotics because of adverse events, patient preference or any other reason.
6) resource allocation using; a) length of inpatient hospital stay b) frequency of outpatient visits c) antibiotic prescribing costs.
7) Quality of life, as evaluated by EQ-5D-3L questionnaire
8) Oxford Hip and Knee Scores (where infection is in the hip or knee)
9) Adherence, as indicated by MEMS (see below) in a subset of participants.

3.2 *Study Design*

The current OVIVA trial is a multi-centre, open label, randomised non-inferiority two-arm pragmatic parallel group clinical trial (one year follow-up), in 1050 people with serious bone and joint infection.

Date of start of recruitment:	26/03/2012– start of main study 03/06/2010 – start of internal pilot
Date of end of recruitment*:	31/10/2015
Date of end follow-up*:	31/10/2016
Date of analysis*:	01/11/16 – 20/01/17
Target number of subjects:	1050 (approximately 525 per arm) including the pilot

*Originally, recruitment to the OVIVA study was to conclude at the end of October 2014. Due to the initial recruitment being lower than expected, the trial was granted a no-cost extension. The above presented timelines take into account this extension.

Participating Centres (NHS Trusts) include:

Oxford University Hospitals NHS Trust; Guy's and St. Thomas' Hospitals NHS Foundation Trust; Royal Free London NHS Foundation Trust; Royal National Orthopaedic Hospital NHS Trust; Birmingham Heart of England NHS Foundation Trust; Royal Liverpool and Broadgreen University Hospitals NHS Trust; Cambridge University Hospitals NHS Foundation Trust; Leeds Teaching Hospital NHS Trust; Sheffield Teaching Hospitals NHS Foundation Trust; University Hospitals Bristol NHS Foundation Trust; Newcastle upon Tyne Hospitals NHS Foundation Trust; Hull and East Yorkshire Hospitals NHS Trust; Brighton and Sussex University Hospitals NHS Trust; Maidstone and Tunbridge Wells NHS Trust; Norfolk and Norwich University Hospitals NHS Foundation Trust; Northampton General Hospital NHS Trust; Northumbria Healthcare NHS Foundation Trust; Queen Elizabeth Hospital King's Lynn NHS Foundation Trust; University Hospital of North Staffordshire NHS Trust; Medway NHS Foundation Trust; Royal Cornwall Hospitals NHS Trust; NHS Tayside; NHS Lothian and NHS Greater Glasgow and Clyde; North West London Hospitals NHS Trust; Blackpool Teaching Hospitals NHS Foundation Trust; Calderdale & Huddersfield NHS Foundation Trust; Royal United Hospital Bath NHS Trust

3.3 *Eligibility*

Inclusion Criteria

1) A clinical syndrome comprising any of the following;
    a) localized pain
    b) localized erythema
    c) temperature >38.0ºC
    d) a discharging sinus or wound
2) willing and able to give informed consent
3) aged 18 years or above
4) the patient has received 7 days or less of intravenous therapy after an appropriate surgical intervention to treat bone or joint infection (regardless of pre-surgical antibiotics) or, if no surgical intervention is required, the patient has received 7 days or less of intravenous therapy after the start of the relevant clinical episode.
5) has a life expectancy > 1 year
6) has a bone and joint infection in one of the following categories;
    a) Native osteomyelitis (i.e., bone infection without metalwork) including haematogenous or contiguous osteomyelitis, and long bone, skull, foot or other foci
    b) Native joint sepsis treated by excision arthroplasty
    c) Prosthetic joint infection treated by debridement and retention, by one stage revision or by excision of the prosthetic joint (with or without planned re-implantation)
    d) Orthopaedic device or bone-graft infection treated by debridement and retention, or by debridement and removal
    e) Spinal infection including discitis, osteomyelitis and/or epidural abscess.

Exclusion Criteria

1) Staphylococcus aureus bacteraemia on presentation or within the last 1 month

2) bacterial endocarditis on presentation or within the last month (NB there are no study mandated investigations. Participants are not required to have echocardiograms, blood cultures, or any other investigations to exclude endocarditis in the absence of a clinical indication)
3) Any other concomitant infection which, in the opinion of the clinician responsible for the patient, required a prolonged intravenous course of antibiotics (e.g. mediastinal infection or central nervous system infection)
4) Mild osteomyelitis, defined as osteomyelitis which, in the opinion of the clinical investigator, would not usually require a 6 week course of intravenous antibiotics
5) An infection for which there are no suitable antibiotic choices to permit randomization between the two arms of the trial (for instance, where organisms are only sensitive to intravenous antibiotics, which occurred in <5% of patients during recruitment for our pilot study)
6) Previous enrolment in the trial
7) Septic shock or systemic features requiring intravenous antibiotics in the opinion of the treating clinician (the patient may be re-evaluated if these features resolve)
8) The patient is unlikely to comply with trial requirements following randomization (including specific requirement for PO or IV course) in the opinion of the investigator
9) There is clinical, histological or microbiological evidence of mycobacterial, fungal, parasitic or viral aetiology
10) The patient is receiving an investigational medical product as part of another clinical trial

The use of antibiotic-loaded cement in spacers or beads at the site of infection will not be an exclusion criterion, but will be recorded in baseline data. Pregnancy, renal failure and liver failure will not be exclusion criteria provided suitable antibiotic choices can be identified.

3.4 *Treatment Interventions*

Eligible patients will be randomized (1:1) to complete the first 6 weeks of antibiotic therapy with the selected course of either IV or PO antibiotic therapy. The selection of individual antibiotics within the allocated strategy (i.e. PO or IV antibiotics) will depend on microbiological assessments, the side effect profile of different antibiotics, patient preferences and epidemiological factors suggesting the likelihood of antibiotic-resistance organisms. Treatment decisions will be left to the clinician caring for the patient, but should remain within the randomized strategy (i.e., either PO or IV antibiotics). If there is no suitable empirical oral antibiotic choice for an individual patient while waiting for culture results, the clinician responsible for the patient may prolong IV antibiotic therapy without withdrawing the patient from the PO antibiotic strategy, provided IV prescribing does not continue beyond 7 days after the beginning of the episode (i.e. after an appropriate surgical procedure or the start of antibiotic prescribing for the clinical episode being treated).

If a participant requires surgery, or experiences an intercurrent illness causing vomiting, inability to swallow, or any other concern about absorption of oral medication, then IV antibiotic therapy may be substituted for a brief period without withdrawing the patient from the randomized strategy. This period should be no longer than 5 days if the patient is to remain “according to protocol”. Note that even if IV antibiotic prescribing exceeds the limits set in the PO strategy, the patient will still contribute to “intention to treat” analysis, and study follow up should therefore continue.

Adjunctive oral antibiotics will be allowed at any stage in the IV group (e.g. oral rifampicin may be added to intravenous antibiotics).

However, if at any point continuing in the randomized strategy (IV or PO) is no longer compatible with good clinical care, the study participant will discontinue the randomized treatment. Study related follow up will continue unless the participant declines this, and the participant will be included in intention to treat analysis. Appropriate reasons for discontinuing the allocated treatment would be that no suitable medication can be selected within the allocated strategy because of adverse reactions, contraindications and susceptibility testing results. Failure to maintain intravenous access is an appropriate reason for discontinuing IV antibiotics and switching to PO antibiotics to complete the first 6 weeks. A wound discharge, superficial erythema or other clinical sign related to infection or resolution of infection is not an appropriate indication for changing PO to IV or vice versa, since there is equipoise regarding efficacy.

If a patient is to be withdrawn from the randomized strategy, this should be discussed with the study CI, the trial physician or another delegate of the CI beforehand. Changing the antibiotic used while remaining within the allocated strategy need not be discussed, but should be done by a clinician with appropriate training in managing infection. Patients who are withdrawn from the allocated strategy should nevertheless continue to be followed up using the trial protocol.

Patients who are withdrawn from their allocated treatment will be included in "intention to treat" analysis of efficacy, but not in the "according to protocol" analysis. Patients who meet a study endpoint may remain in the PO strategy for purposes of selecting their ongoing antibiotic treatment, since there is equipoise regarding the relative efficacy of PO and IV antibiotic treatment.

Dose adjustments based on renal or hepatic function, drug interactions or other factors will be made by the clinician according to drug labelling information, the British National Formulary and local pharmacy guidelines.

The dose and antibiotics used will be recorded in the CRF at scheduled reviews.

3.5 *Sample Size*

Original sample size calculation:

In the Oxford pilot, 10 participants experienced a primary endpoint among the first *197* randomizations. Based on a 5% event rate, we will require 950 evaluable participants for sufficient power (at one-sided alpha=0.05 and power=90%) to determine that the PO strategy is non-inferior to the IV strategy, defined as the upper 90% confidence limit for the difference being less than a 5% absolute increase in event rate (i.e. an increase to 10%). To compensate for participants being lost to follow up (allow for approximately 10%), and to ensure that the "according to protocol" analysis retains reasonable power, we will aim to recruit 1050 participants.

Updated sample size calculation:

After the interim analysis, the sample size calculation for the OVIVA trial was updated as follows:

In the Oxford pilot, 10 participants experienced a primary endpoint among the first *197* randomizations. Based on an anticipated 5% event rate, we estimated that 950 evaluable participants (uplifted to 1050 to account for loss to follow up and to allow for per protocol analyses) would be necessary (at one-sided alpha=0.05 and power=90%) to determine that the PO strategy is non-inferior to the IV strategy, defined as the upper 90% confidence limit for the difference being less than a 5% absolute increase in event rate (i.e. a relative increase of 100%). Following an interim analysis in March 2015, pooled data from the multicentre trial over a 1 year follow-up period demonstrated that the true event rate is plausibly closer to 12.5%. In response to this finding, we have adjusted the non-inferiority margin to 7.5% (i.e. a relative increase of 60%) with explicit agreement from the DMC. Using 90% power and a one-sided alpha of 0.05, a minimum of 744

participants would be required, allowing for a 10% loss to follow-up. As the final control group failure rate remains unknown, and to optimise the potential utility of subgroup analyses, the recruitment target will remain 1050.

3.6 *Strategies for achieving adequate recruitment*

During the trial, regular telephone conferences and a trial specific website were implemented to enable sites to share good practice and to allow for discussion around recruitment rates and protocol adherence. In addition, the trial has been publicised and additional sites have been included. Monthly updates of recruitment numbers by site are circulated and personal contact with PIs and their research teams are maintained where necessary.

3.7 *Randomisation*

Trial participants will be randomised (1:1) to either the PO or IV treatment strategy using a randomisation list with varying block sizes stratified by site.
The randomisation schedule, consisting of one list per site, will be prepared by the trial statistician and transferred to the OCTO programming team using secure methods of transfer. The lists will be held securely by the trial statistician and the OCTO programming team. OCTO will provide the randomisation database and randomisation services support.
The trial statistician conducts regular checks to ensure the randomisation is working as expected.

3.8 *Hypotheses and Definition of Primary and Secondary Outcomes*

Primary endpoint:
The primary endpoint of the OVIVA study is definite failure of infection treatment identified within 12 months from randomisation, whereby definite failure is indicated by one or more of the following:
a) isolating bacteria from 2 or more samples of bone/spine/peri-prosthetic tissue, where the bacteria are similarly typed
b) a pathogenic organism (e.g. *Staphylococcus aureus* but not *Staphylococcus epidermidis*) on a single, closed, biopsy of native bone or spine
c) diagnostic histology on bone/peri-prosthetic tissue
d) formation of a draining sinus tract arising from bone/prosthesis or
e) recurrence of frank pus adjacent to bone/ prosthesis.

* “similarly typed” refers to the results of routine laboratory work, including bacterial genus/species and the results of routine antibiotic susceptibility testing. We will not require any additional bacterial typing in the laboratory beyond local routine practice.

H_o: The proportion of participants with a definitive treatment failure in the PO group is more than 7.5% higher than the proportion of participants with definitive treatment failure in the IV group:
$p_{PO} - p_{IV} > 7.5\%$, where p_{PO} and p_{IV} are the proportions of participants with definitive treatment failures randomised to the PO and IV strategies respectively
H_1: The proportion of participants with a definitive treatment failure in the PO group is not more than 7.5% higher than the proportion of participants with definitive treatment failure in the IV group:
$p_{PO} - p_{IV} <= 7.5\%$, where p_{PO} and p_{IV} are the proportions of participants with definitive treatment failures randomised to the PO and IV strategies respectively

Secondary endpoints:
All statistical tests for the secondary endpoints are standard two-sided superiority tests with the exception of 4) below, which is analysed using a non-inferiority approach in line with the primary endpoint.

1) SAEs, including death (i.e. all cause) according to treatment allocation.
   H_o: There is no difference in the odds of experiencing at least one SAE in both randomised trial arms:
   $OR_{PO/IV} = 1$, where $OR_{PO/IV}$ = odds of experiencing an SAE in the PO arm / odds of experiencing an SAE in the IV arm
   H_1: There is a difference in the odds of experiencing at least one SAE between the randomised trial arms:
   $OR_{PO/IV} \neq 1$, where $OR_{PO/IV}$ = odds of experiencing an SAE in the PO arm / odds of experiencing an SAE in the IV arm
2) The frequency of line complications (i.e. infection, thrombosis or other events requiring early removal or replacement of the line).
   As this summary includes primarily participants randomised to the IV strategy, no formal statistical tests will be performed.
3) The proportion of participants with *Clostridium difficile* associated diarrhoea in each treatment arm.
   H_o: There is no difference in the odds of experiencing at least one with *Clostridium difficile* associated diarrhoea in both randomised trial arms:
   $OR_{PO/IV} = 1$, where $OR_{PO/IV}$ = odds of experiencing *Clostridium difficile* associated diarrhoea in the PO arm / odds of experiencing *Clostridium difficile* associated diarrhoea in the IV arm
   H_1: There is a difference in the odds of experiencing with *Clostridium difficile* associated diarrhoea between the randomised trial arms:
   $OR_{PO/IV} \neq 1$, where $OR_{PO/IV}$ = odds of experiencing with *Clostridium difficile* associated diarrhoea in the PO arm / odds of experiencing with *Clostridium difficile* associated diarrhoea in the IV arm
4) The frequency of the secondary endpoints "probable" or "possible" treatment failure as composites with definitive treatment failure. These will be determined by blinded endpoint committee review, and determined according to the following criteria;
   a) Loosening of a prosthesis, confirmed radiologically OR
   b) non-union of a fracture after 6 months, confirmed radiologically OR
   c) superficial spreading erythema, treated as cellulitis with an antibiotic for >1 week; where results from deep tissue samples do not meet the primary endpoint as described above.
   Where appropriate deep tissue samples are sent for microbiology and results of culture are negative, either of a), b) or c) are met, then the endpoint will be regarded as "possible". On the other hand, where deep tissue samples are not sent for microbiology, and either a), b) or c) are met, then the endpoint will be regarded as "probable".

   H_o: The proportion of participants with any treatment failure in the PO group is more than 7.5% higher than the proportion of participants with any treatment failures in the IV group.
   $p_{PO} - p_{IV} > 7.5\%$, where p_{PO} and p_{IV} are the proportions of participants with any treatment failures randomised to the PO and IV strategies respectively
   H_1: The proportion of participants with any treatment failure in the PO group is not more than 7.5% higher than the proportion of participants with any treatment failure in the IV group.
   $p_{PO} - p_{IV} <= 7.5\%$, where p_{PO} and p_{IV} are the proportions of participants with any treatment failures randomised to the PO and IV strategies respectively

5) Early termination of the planned 6 week period of oral or IV antibiotics because of adverse events, patient preference or any other reason.

H_o: There is no association between early termination of the planned six week strategy and the randomisation allocation.

H_1: There is an association between early termination of the planned six week strategy and the randomisation allocation.

6) Resource allocation determined by; a) length of inpatient hospital stay b) frequency of outpatient visits c) antibiotic prescribing costs.
   Refer to the separate health economics analysis plan for the hypotheses for the relevant analyses.
7) Quality of life evaluated by EQ-5D-3L questionnaire
   H_o: There is no difference in the median EQ-5D-3L index between the two randomised trial arms
   Median (EQ-5D-3L $_{PO}$) = Median (EQ-5D-3L $_{IV}$)
   H_1: There is a difference in the median EQ-5D index between the two randomised trial arms
   Median (EQ-5D-3L $_{PO}$ ≠ Median (EQ-5D-3L $_{IV}$)
8) Oxford Hip and Knee Scores (where infection is in the hip or knee)
   H_o: There is no difference in the median OHS/ OKS between the two randomised trial arms.
   Median (OHS_{PO} = Median (OHS_{IV})
   Median (OKS_{PO} = Median (OKS_{IV})
   H_1: There is a difference in the median OHS/ OKS index between the two randomised trial arms.
   Median (OHS_{PO} ≠ Median (OHS_{IV})
   Median (OKS_{PO} ≠ Median (OKS_{IV})
9) Adherence to oral medication in terms of the MEMS caps. As this summary includes participants randomised to the PO strategy only, no formal statistical tests will be performed.

Secondary endpoints 1, 2, 4 and 5 will be determined by study clinicians. The primary endpoint and secondary endpoint 4 will be determined by the blinded endpoint committee using redacted notes. Secondary endpoints 6 and 7 will be determined by participants with evidence of infection in the hip and knee respectively using questionnaires. Secondary endpoint 8 will be determined in a subset (i.e. Oxford, Guy's and St Thomas' Trusts, Royal Free Hospital Trust and the Royal National Orthopaedic Hospital) using MEMS.

3.9 *Outcomes Assessment Schedule*

Baseline assessments are performed prior to randomisation on day 0.
Table 1 below details all important time points and assessments in the study.

Table 1: OVIVA assessment schedule

Time	Activity
Day -7 to 0	Definitive surgical procedure (see above for definition) or, where not applicable, the start of antibiotic treatment for the current clinical episode of illness should be within this period.
Antibiotic prescribing	
Day 0	Randomized to oral vs IV strategy. May continue on intravenous antibiotics within the "oral strategy" up to 7 days in total (including pre-randomization IV antibiotics given for current clinical episode).
Days 0-42	Period during which randomized therapy (i.e. Oral or intravenous antibiotics) is given. MEMS will be provided if applicable (see below)
Day 42 onwards	May receive further oral antibiotics as clinically appropriate. These further antibiotics are not determined by randomization.
Clinic Reviews	
Day 42 (accepted range 21 to 63)	Investigator completes 1st review. Collects MEMS if used.
Day 120 (accepted range 70 to 180)	Investigator completes 2nd review. Collects MEMS if used and not previously collected.
Day 365 (accepted range 250 to 420)	Investigator completes 3rd review and end of study follow up.
Questionnaires	
Day 0, 14, 42, 120, 365 and at endpoint or SAE	EQ-5D-3L questionnaire
Day 0, 120, 365	Oxford Hip/Knee Questionnaire

3.10 *Data Management Responsibility*

Monitoring involves overseeing the progress of the trial by confirming the data is accurate, complete and verifiable from source documents. Using the OVIVA Monitoring Plan V1, Sept 2014, we are conducting monitoring visits to our collaborating sites, which involves confirmation of correct consenting and storage, reviewing of eligibility before randomisation, primary outcome data, CRF validation, questionnaire data accuracy against source data, and safe storage of all data and documentation. Using the OpenClinica Database, the study co-ordinator regularly reviews any missing data, and sends sites data missing reports using the OVIVA Data Queries/Monitoring Form V1, Sept.2014 (adapted from OCTRU-OF-015_V1.0).

4. Quality Control and Data Validation

Throughout the trial, data checks will be performed in conjunction with data collection and data entry.
Prior to any analysis, the Trial Statistician will perform additional data checks and validations, investigating the data for outliers and inconsistent dates. All apparent outliers will be checked against paper records and either confirmed as valid observations or corrected.
For the final analysis a manual 100% data entry check of the results of the reviews performed by the Endpoint Review Committee against the information on treatment failures as read into Stata will be performed. The results from the review are usually received in table format (e.g. Microsoft Excel). This review will include all participants for whom potential treatment failures have been recorded and whose redacted notes have therefore been reviewed by the Endpoint Review Committee.
Data entry for PROMS (i.e. the EQ-5D-3L, the compliance questionnaire for PO patients and the OKS/ OHS where appropriate), as well as baseline infection categories as defined by the endpoint review committee (for non-definite infections) will be checked against the paper CRFs for 20 patients. Additional data checks are performed if the error rate is found to be greater than 1%. Using the OVIVA Study Monitoring Plan (V1, Sept 2014), we have commenced checking the baseline infection rates, and all questionnaire data against source data in the clinical notes and from microbiology results, and from source questionnaires for 10% of the total study participants, for two collaborator sites, so far. We intend to continue with more monitoring visits over the next few months. The OpenClinica database is regularly checked and queries are raised with collaborating sites for possible inconsistencies and missing data (see 3.2)

The analysis for the primary endpoint will be repeated by a second statistician. The performance of a second analysis for the primary endpoint will be reported in the final statistical report. Information on randomisation allocation and endpoints will be cleaned and transferred securely to the second statistician, who will independently perform the primary outcome analysis in Stata, or another validated statistical package.

The statistical report will be reviewed by a second statistician to ensure that the SAP/principles of the SAP have been followed as per the OCTRU SOP STATS-005.

5. Data SAFETY monitoring Committee and Interim Analyses

A data safety monitoring board will be formed, which is independent from the study team and the sponsor. The DMC will be composed of 3 members; Neil French (chair, Professor of Infectious Disease, Liverpool University), Colette Smith (Lecturer in Biostatistics, UCL) and Martin Llewelyn (Reader in Infectious Diseases and Therapeutics, Brighton and Sussex University). If, during the course of the trial, one of the DMC members withdraws, a replacement with a similar background will be identified.

The DMC will meet (either in person or by teleconference) to discuss the study design and SOPs shortly before the start of the study. Investigators will participate in this meeting. The DMC will also evaluate the frequency of endpoints in an unblinded analysis, when investigators will not be present. The DMC will make a recommendation before investigators proceed with the multi-centre trial.
A full interim analysis including all available data from all sites will be reviewed by the DMC after approximately 100 participants from sites other than Oxford have been recruited and completed their follow-up to review the safety and ethics of the OVIVA trial.
Extra meetings may be convened at the request of the investigators, sponsor, or DMC members to discuss emerging data that is a cause for concern.

It is expected that the DMC would only recommend early stopping if there was a very significantly worse outcome in the PO antibiotic group compared to the IV group (i.e. using the Haybittle-Peto stopping boundary).
The DMC will discuss the analysis plan before the investigators conduct the final analysis

6. Descriptive Analyses

6.1 *Representativeness of Study Sample and Patient Throughput*

A complete CONSORT flow diagram will be included in the trial report, clearly stating the number of patients screened, eligible, randomised and followed-up throughout the trial. Information on reasons for ineligibility will be given; information on randomisations and follow-up will be presented by treatment arm and detail how many participants received their allocated intervention.

6.2 *Baseline Comparability of Randomised Groups*

For all information collected at baseline, numbers (with percentages) for binary and categorical variables (including gender) and means (with standard deviations), or medians (with the interquartile range and range) for continuous variables (including baseline patient reported outcomes and age) will be presented overall and by treatment group.
There will be no tests of statistical significance or confidence intervals for differences between randomised groups on any baseline variable because, by definition of randomisation, these arise only due to chance.

6.3 *Comparison of Losses to Follow-up*

The numbers (with percentages) of losses to follow-up (defaulters and withdrawals) over the one year period of the study will be reported and compared between the PO and IV groups using frequency and percentages. Any deaths (and their causes) will be reported separately within the section on SAEs and complications.

6.4 *Description of Available Data*

The availability of data for baseline assessments as well as for primary and secondary endpoints will be described for all appropriate trial time points.
Data items are defined as available if either the clinic assessment form has been completed, or for patient reported outcome measures, if the information provided can be used in the analysis. For example, the OKS/ OHS final scores can only be calculated when no more than two items are missing. Hence the OKS/ OHS will be classed as available if the responses to at least 10 of the 12 items are available.

Summaries will be provided overall and by trial arm, and the number of available data items will be presented together with the number of data item expected and a percentage indicating the rate of data compliance for each endpoint and time point (i.e. investigating what percentage of expected data is actually available).

6.5 *Description of Compliance with Intervention*

Early termination of the planned six week period of oral or IV antibiotics, as well as adherence to the medication are secondary endpoints of the OVIVA trial and will be summarised in the endpoint relevant section.

6.6 *Unblinding of Randomised Treatments*

N/A – OVIVA is an open label trial and participants and staff are not blinded to treatment allocations, but the independent Endpoint Review Committee is blinded to participants' treatment allocations.

6.7 *Reliability*

The trial is open-label, as blinding is not possible, since giving a prolonged intravenous placebo treatment was considered unethical. Open label studies are at risk of bias. Objective criteria for meeting the primary endpoint were therefore set out, which will be examined by a blinded endpoint review committee.

For any participant that is admitted to hospital with signs or symptoms relating to the original site of infection, investigators will send a redacted copy of the inpatient admission notes to the endpoint review committee. Notes will be redacted for personal identifiable information and for antibiotic names or routes of administration. One member of the committee will be expected to review the notes in detail, and summarise the key findings that determine an endpoint for the other committee members. Blind to the treatment allocation, the committee will determine an endpoint either by consensus following discussion, or by a vote called by the chair if consensus cannot be reached.
The endpoint committee will meet at regular intervals throughout the recruitment and follow-up of the trial, to ensure that up-to-date information on endpoints is available for interim DMC meetings.

With regards to the trial outcomes, the endpoint committee will only be required to review potential treatment failure. All other secondary endpoints including SAEs, line complications, early termination of treatment patient reported outcome data and data for resource allocation will be determined directly by the local study clinicians, or completed by the trial participants.

The endpoint committee will also have a role in determining diagnostic sub-groups for the infection criteria at baseline, following the guidance listed below:

"Definitive" evidence of infection, defined by one or more of the following:

a) isolating bacteria from 2 or more samples of bone/spine/peri-prosthetic tissue, where the bacteria are similarly typed
b) a pathogenic organism (e.g. *Staphylococcus aureus* but not *Staphylococcus epidermidis*) on a single, closed, biopsy of native bone or spine
c) diagnostic histology on bone/peri-prosthetic tissue
d) a draining sinus tract arising from bone/prosthesis or
e) frank pus adjacent to bone/ prosthesis.

If any of these criteria are met, then the category "definitive" infection will be applied without endpoint committee review.
Where these criteria are not met, the endpoint committee will be sent a redacted copy of the patient's admission notes and laboratory results from the time of randomisation, and apply the following criteria to determine "probable" or "possible" infection:

Infection will be categorized as "probable" where microbiological sampling has not been undertaken, AND none of the other criteria for definite infection are fulfilled AND any one of the following are met:

a) Radiological or operative findings of periosteal changes suggesting chronic osteomyelitis OR
b) Radiological findings suggesting discitis/spinal infection OR
c) The development of a discharging wound after an orthopaedic procedure where prosthetic material has been implanted OR
d) The presence of deep pus close to but not adjacent to bone/prosthetic joint/orthopaedic device OR
e) The presence of peri-prosthetic necrotic bone OR
f) Rapid loosening of a joint prosthesis/orthopaedic device (i.e. leading to localized pain in less than 3 months since implantation) in the absence of a mechanical explanation for rapid loosening.

Infection will be categorized as "possible" where microbiological sampling has been undertaken with negative results (according to criteria described above for "definite" infection) AND other criteria for definite infection are not fulfilled AND in addition one or more of the criteria listed a) to e) above is met.

A sample of all derived and generated variables to be used for the trial analysis will be verified, in accordance with the OCTRU SOP STATS-003.

7. Patient Groups for Analysis

The following patient populations will be utilised in the analyses:

Intent to treat (ITT): All randomised participants will be analysed according to their allocated intervention.

Modified intention to treat analysis (MITT): Randomisation participants will be analysed according to their allocated intervention if they have non-missing outcome data. For adjusted analyses, relevant baseline variables that are used to adjust the model also need to be available in order for participants to be included in the MITT population.

Per protocol (PP): All participants who have received at least four weeks of their randomised strategy, and, if in the PO group, did not exceed the limits set for the use of IV antibiotics (i.e. 5 days continuously at any one time). Participants who were recorded to have exited early from their randomised strategy due to possible or probable recurrence of infection will also be included in the PP population. Participants will be included in the PP analyses if sufficient outcome and baseline data (where relevant) is available.

8. Analyses to address primary aims

It is anticipated that the analysis will use STATA statistical software, or other validated statistical software, such as SAS or R (versions will be recorded in the Statistical report).

8.1 *Evaluation/Definition of Primary Outcome (where applicable)*

The primary endpoint of the OVIVA trial, i.e. definite failure of infection treatment, as defined in section 3.8, is reached if any of the reports of potential treatment failures as recorded by the local

clinical team are confirmed as a definite failure of infection treatment by the endpoint review committee. This endpoint will be analysed primarily as a binary outcome (i.e. not as a time to event outcome) because dates may reflect timing of observations rather than actual failure.

8.2 *Statistical Methods Used for Analysis of Primary Outcome*

Primary analysis

Based on the intention to treat population, the proportions of participants experiencing the primary endpoint (i.e. definitive treatment failure as adjudicated by a blinded endpoint review committee) will be tabulated by treatment group (i.e. oral vs intravenous therapy). If the absolute, upper two-sided 90% confidence interval (CI) around the absolute unadjusted difference (i.e. oral-intravenous) is less than 7.5%, then the criteria of non-inferiority will be met.

The primary analysis is an unadjusted analysis. Therefore, a complete cases analysis, whereby participants with missing outcome data are excluded, makes the assumption that the data is missing completely at random. This is, the probability of data being missing does not depend on observed or unobserved measurements. This is a very strong assumption, which is unlikely to hold in practice.

Therefore, the ITT population forms the basis of the primary analysis. This includes all randomised participants within their randomised treatment allocations regardless of their compliance with the protocol. Participants with missing outcome data are not excluded from this analysis. Therefore assumptions have to be made about their outcomes.

The originally specified analysis classed individuals with incomplete follow-up and no event observed to date as not having experienced an endpoint. This is essentially a single "hard" imputation of no event for these participants. This analysis will now be performed as a supporting analysis, and multiple imputation (MI) will form the basis of the primary analysis.

Under MI, data are assumed to be missing at random, i.e. missing data are dependent on the values of observed data, but are independent of the values of the missing data themselves once observed data have been accounted for. This assumption is more likely to hold in practice than the missing completely at random assumption, and its robustness can be assessed in appropriate sensitivity analyses.

MI imputes missing data based on information from other observed variables. Several imputations are generated and combined under Rubin's Rule to account for the uncertainty around the imputed values(2). Missing values for the primary outcome will be imputed based on a logistic regression model, such as the *mi impute* command in Stata.

Hence, the primary analysis of the OVIVA trial is based on the ITT population whereby missing data is handled using an MI approach.

The following variables are used in the imputation model, and were identified as relevant in predicting outcomes by the OVIVA CI:

- infection details at baseline are combined as follows and used as binary variables in the imputation model:
  - Chronic osteomyelitis debrided, no current implant or device OR discitis/ spinal osteomyelitis/ epidural abscess debrided
  - Chronic osteomyelitis as above, but not debrided OR discitis/ spinal osteomyelitis/ epidural abscess but not debrided
  - Implant or device present and retained ("DAIR")
  - Removal of orthopaedic device for infection OR prosthetic joint implant removed
  - Prosthetic joint implant, 1-stage revision
- Whether or not antibiotic beads/ cement were used in the index operation
- Participants' comorbidity status (yes vs. no):
  - Diabetes

    - Peripheral vascular disease in participants with foot infections
    - Current smoker
    - Rheumatoid arthritis or systemic autoimmune disease
- Staph Aureus present in samples taken before randomisation
- Pseudomonas sp present in samples taken before randomisation
- Age
- Gender

Due to the large number of binary variables used in the MICE model, resulting in a high likelihood of perfect predictions, convergence issues of the imputation model are anticipated. This will be addressed by augmenting the data, i.e. adding a small number of additional observations with small weights when model parameters are estimated to prevent perfect prediction(3, 4).
Non-linearity in the relationship between age and outcome will be explored in the complete cases. If there is clear evidence of non-linearity, the multiple imputation model will be adjusted appropriately (for example, age may be modelled using natural cubic splines).

Supporting analyses

A number of supporting analyses will be performed. These will focus on the consistency of the point estimates and two-sided 90% CIs rather than formal comparison with the 7.5% non-inferiority margin. Details of these analyses are given below, or in the section on subgroup analyses:

Initial supporting analyses will include the following deviations from the above described primary analysis, using different analysis populations and assumptions about missing data about:

- The MITT population will be used, i.e. the analysis will be performed on the complete cases only, without imputation of missing outcomes. Participants are analysed based on their randomisation allocation.
- The ITT population will be used; however, in this analysis, all participants with incomplete follow-up and no event observed to date will be classed as not having experienced an endpoint (single imputation). Death without clinical failure is not classed as a treatment failure for this analysis. This analysis was initially defined as the primary trial analysis, but was moved to the supporting analyses in favour of a multiple imputation approach for handling missing data. Participants are analysed based on their randomisation allocation.
- The PP population will be used. Participants are analysed based on their randomisation allocation, but are excluded from the analysis if they do not meet the PP population criteria.

In addition, a logistic regression model will be used to calculate the estimates of the treatment differences for the occurrence of definite treatment failure as adjudicated by the blinded endpoint review committee adjusted for age, comorbidity, infecting pathogen, and type of infection.
Additional information on the categorisation of the infecting pathogen and type of infection can be found in sections 0 and 0 respectively. Categories with low counts may be combined.
Information on 11 comorbidities is collected at enrolment, and these comorbidities will be added to the model as separate binary variables. In the event of comorbidities with very low counts, these comorbidities may be combined to avoid difficulties with the maximum likelihood estimation of the logistic model. Where no information has been entered on the comorbidities, the participants will be considered not to suffer from these comorbidities. The imputed endpoints and explanatory variables from the primary analysis will be used; however, participants with missing data for the infecting pathogen will be excluded from this analysis.

For the multivariate logistic regression models, residual and predicted values produced form the model will be examined to assess the assumptions of the model. Specifically, the assumption of linearity between the predicted log odds and the covariates is assessed by plotting lowess graphs. The independence of the error terms will be considered. Influential cases are investigated by plotting the standardised Pearson's residuals against the predicted probabilities and the leverage of the individual observations.

To assess any potential bias in the post-randomisation surveillance, which would present as a delay in time to meeting a definitive endpoint in one randomised group, as well as loss to follow-up or death without and event, a time to event analysis will be performed.
The Cox proportional hazards model (if appropriate) will be used to compare the time to first treatment failure between the trial arms. The model will not be adjusted for baseline characteristics, as this analysis is focussing on the timing of events. Participants with no treatment failures will be censored at the earliest of the following dates: death, last assessment if they are not known to have died and were lost to follow-up prior to their one year assessment, or at the date of their one year follow-up. Treatment estimates, standard errors, hazard ratios and 95% confidence intervals, as well as p-values will be presented. Failure free time to event curves will be calculated using the Kaplan-Meier curves will be presented for the time to meeting an endpoint by trial arm. This analysis will be performed for the ITT population only.

The proportional hazards assumption will be assessed by plotting the hazards over time (i.e. the log cumulative hazard plot) for both treatment arms, investigating the log-log plots of the hazards and a test for proportionality. Should these assessments indicate non-proportional hazard rates, alternative approaches will be examined, e.g. piecewise hazards.

8.3 *Adjustment of P values for Multiple Testing*

There is no multiple testing as only a single primary outcome is considered. All additional analyses are undertaken with an intention to further inform the results from the primary analysis. Therefore significance levels used will be 0.05 and 95% confidence intervals will be reported.

The DMC will review interim summaries and a formal interim analysis. However it is expected that the DMC would only recommend early stopping if there was a very significantly worse outcome in the PO antibiotic group compared to the IV group (i.e. using the Haybittle-Peto stopping boundary). Therefore, the significance level used to determine early termination of the trial is very low (i.e. 0.001) and no formal adjustment of the p-value for the final analysis is considered necessary.

8.4 *Missing Data*

The primary outcome of the OVIVA trial, i.e. definitive treatment failure as adjudicated by a blinded endpoint review committee does not rely on trial specific clinic assessments or patients reports, but can be obtained from hospital notes. Therefore, only minimal amounts of missing data are expected, primarily in cases where participants formally withdraw from all further follow-up or relocate or their medical records can no longer be accessed.

In the primary analysis, multiple imputation is utilised. Additional complete cases analyses are also performed. These analysis make strong assumptions about the underlying missing data mechanism, assuming that data is either missing at random or missing completely at random.

Sensitivity analyses will assess the robustness of these analyses, by also considering the impact on the study results if data are assumed to be missing not at random, i.e. if those with missing data have better or worse outcomes than those with completely observed outcome data. The sensitivity analysis will include a tipping point analysis(5-7), whereby the departures from the missing completely at random assumption needed to change the trial results will be explored. In discussion with the CI and clinical team, the robustness of the trial results with regards to missing data will be discussed.

8.5 *Pre-specified Subgroup Analysis*

All subgroup analyses will be based on the MITT population (complete cases analysis) and presented as forest plots.

8.5.1 Pre- specified Subgroup Analysis considering infection subgroups at randomisation

Taking into account the subgroups of participants with firstly a “definite” infection (vs. “probably”/ “possible” infection) at randomisation, and secondly the participants with a “definite” or “probable” infection (vs. “possible” infection) at randomisation. For the ITT population, a logistic regression model will be constructed with the occurrence of the primary endpoints (i.e. definite treatment failure as adjudicated by the blinded endpoint review committee) as the outcome, and the randomised treatment as well as the subgroups of infection at randomisation (“definite” vs. “probable”/ “possible” infection in the first statistical model, and “definite”/ “probable” vs.

"possible" infection in the second statistical model) as explanatory variables, as well as the interactions between the randomised treatment and the infection subgroup at randomisation.

Note: There are some participants for whom the infection subgroup at baseline could not be confirmed by the review committee. A decision was made by the trial team to include these participants into the "possible infection" category. This is because they were felt to have clinical evidence of infection at randomisation.

8.5.2 Pre-specified Subgroup Analysis considering the type of infection

Sub-group analysis will be used to determine the consistency of treatment effects by type of infection.

Information on the type of infection is collected at the enrolment of trial participants, and categorised as follows:

1. Chronic osteomyelitis debrided, no current implant or device OR
   Discitis/spinal osteomyelitis/ epidural abscess debrided
2. Chronic osteomyelitis as above, but not debrided OR
   Discitis/spinal osteomyelitis/ epidural abscess but not debrided
3. Implant or device present and retained (i.e. "DAIR")
4. Removal of orthopaedic device for infection OR
   Prosthetic joint implant removed
5. Prosthetic joint implant, 1-stage revision
6. OVIVA infection criteria not met

Where participants fall into more than one category, they will be assigned to the highest numeric category in the above list. Categories with very low counts may be combined with the next (lower) category.

For the ITT population, a logistic regression model will be constructed with the occurrence of the primary endpoint (i.e. definite treatment failure as adjudicated by the blinded endpoint review committee) as the outcome, and the randomised treatment as well as the infection type (as a 6 level categorical variable) and the interaction between randomised treatment and infection type as explanatory variables. The test for heterogeneity is the 5df test that the effect of randomised treatment is the same across all levels of infection type, i.e. that each interaction coefficient is zero.

8.5.3 Pre-specified Subgroup Analysis considering the infecting pathogen

Sub-group analysis will be used to determine the consistency of treatment effects by infecting pathogen.

Information on the following five infecting pathogens is collected:

1. Staph Aureus
2. Pseudomonas spp
3. Gram negative organism(s)
4. Streptococcus
5. Coagulase negative Staphylococcus
6. No infecting pathogen present

Where evidence for more than one of the above pathogens is present on the deep tissue microbiology results taken prior to randomisation, they will be assigned to the highest numeric category in the above list. The infecting pathogen will be a single variable with six levels.
The above categories for the infecting pathogens have been chosen as part of a pragmatic approach and include the main gram positive categories. It was felt that insufficient numbers of patients would be available for other infecting pathogens to enable meaningful statistical subgroup analysis.

For the ITT population, a logistic regression model will be constructed with the occurrence of the primary endpoints (i.e. definite treatment failure as adjudicated by the blinded endpoint review committee) as the outcome, and the randomised treatment as well as the infecting pathogen and the interaction between randomised treatment and infecting pathogen.

8.5.4 Pre-specified Subgroup Analysis considering the intended and actual antibiotic choice

In some centres, randomisation to oral antibiotics may result in an increased use of antibiotics with particular properties in penetrating biofilms, such as rifampicin. Subgroup analysis will be used to assess the effect of potentially different treatment choices between the trial arms.

Both intended IV and oral antibiotic choices pre-randomisation, and actual antibiotic choices post-randomisation to either oral or IV, were collected. Actual antibiotic choices are a post-randomisation variable and therefore it is not possible to exclude some influence of randomisation on these choices. This will be assessed by comparing intended vs. actual antibiotics for the group the patient was actually randomised to.

As there is particular interest in rifampicin, a specific subgroup analysis will be conducted for this variable. A variable will be created indicating whether or not rifampicin was an antibiotic choice for the intravenous and oral arm, using the treatment intentions for both treatments as recorded prior to randomisation.

Using the the ITT population, a logistic regression model will be constructed with the occurrence of the primary endpoints (i.e. definite treatment failure as adjudicated by the blinded endpoint review committee) as the outcome, and the randomised treatment as well as the above described indicator variable (rifampicin was an intended treatment option yes vs. no) and the interaction between the two variables.

An additional subgroup analysis will consider the clinician's specific antibiotic intentions recorded prior to randomisation, as a categorical variable. The antibiotic intentions will be categorised into the following groups based on the intended drug. Where multiple antibiotics were taken, patients will be assigned to the highest numeric category in the below list.

Planned IV treatments	**Planned PO treatments**
1. Glycopeptides (i.e. teicoplanin / vancomycin)	1. Penicillins
2. Penicillins	2.Quinolones
3. Cephalosporins	3. Tetracyclines
4. Carbapenems	4. Macrolides / Lincosamide
5. Other single IV antibiotic	5. Other single PO antibiotic
6. Combination IV antibiotics	6. Combination PO antibiotics

For the ITT population, a logistic regression model will be constructed with the occurrence of the primary endpoints (i.e. definite treatment failure as adjudicated by the blinded endpoint review committee) as the outcome, and the randomised treatment as well as the subcategory of the antibiotic intention and the interaction between the two variables.

For all pre-specified subgroup analyses, diagnostic checks will be performed as described in section 8.2.

8.6 *Treatment by Centre Interaction*

Consistency of potential effects will be assessed across all centres by informal examination of the within centre effects. There will be limited capacity to investigate these formally and it is noted that such centre effects are expected by chance.

Treatment allocation by centre interaction will be explored and odds ratios will be presented as forest plots without the performance of statistical tests.

This summary will only include centres where patients in both arms have experience treatment failures, as the odds ratios can otherwise not be estimated.

8.7 *Sensitivity Analysis*

No sensitivity analysis in addition to that discussed in the above sections is planned in the context of the primary analysis. The trial team feels that the above described analyses (including the PP analysis, which is part of the primary analysis described above, and sensitivity analysis to explore the potential effects of missing data) are sufficient to assess the robustness of the trial results.

9. Analysis to address secondary aims

The secondary aims of the study are to determine the effect of oral versus intravenous antibiotic strategies on SAEs, the frequency of line complications, “possible” and “probable” treatment failures as composites with “definite” treatment failures, early termination of the planned six week treatment period, quality of life measured by the EQ-5D-3L for all participants and the OKS/ OHS in the relevant subset of participants, adherence to the allocated intervention and cost-effectiveness. These analyses are performed on the MITT population.

More details on the secondary endpoints are provided in section 141.

9.1 *Evaluation/Definition of Secondary Outcomes (where applicable)*

- The “probable” and “possible”, as well as “definite” treatment failures are determined by the blinded endpoint review committee and are not derived as part of this analysis.
- Early termination of the planned 6 week period of oral or IV antibiotics because of adverse events, patient preference or any other reason is defined as exiting the allocated strategy
- The patient reported outcomes (EQ-5D-3L, OKS and OHS).

  For MEMS, adherence will be calculated by the supplies, medAmigo, as follows: During the period of monitoring, a day-by-day proportion of correct dosing is calculated by dividing the number of MEMS openings by the number of dose prescribed that day. When there are more MEMS openings than dose prescribed that day, these extra openings (can be driven by extra intakes or artificial openings for a refill/data download) are not taken into account in the calculation. This implies that the calculation is capped by 100% or overdose is not taken into account.

9.2 *Statistical Methods Used for Analysis of Secondary Outcomes*

9.2.1 "possible" and "probable" treatment failures as composites with "definite" treatment failures

A breakdown of the types of treatment failures recorded by trial arm will be provided, together with a summary of the number and type of treatment failures experience within each arm.
The primary analysis described in section 8.2 will be repeated for occurrence of the composite of "possible", "probable" and "definite" treatment failures. Secondary analyses described in section 8.2 will not be performed. Subgroup analyses described in section 8.5 will be performed for the MITT population only.

9.2.2 Adverse events and complications

Episodes of *Clostridium difficile* will be summarised overall and by treatment arm (frequency and percentages). Participants will be categorised as either having or not having experienced episodes of *C. difficile*. Using this as a binary outcome variable, the unadjusted risk differences in episodes of Clostridium difficile between the treatment arms will be reported for the MITT population.

Reported serious adverse events will also be presented in this section. This includes the number of participants with at least one recorded severe adverse event, as well as the number of severe adverse events reported per participant. In addition, summaries will include the timing of the report from randomisation and whether complications were expected and/ or thought to be related to the randomisation, and the outcome of any SAEs will be summarised. Full details will be given for SAE that are related to the randomisation.

A Chi-squared test will be used to assess if there is evidence of an association between the allocated treatment and the occurrence of at least one SAE for participants (using a binary indicator variable).

9.2.3 The frequency of line complications

Details of the IV lines used in each arm of the trials will be summarised, detailing the frequency of percentage of PIC, Hickman and other lines used.
The number of participants with line complications on each arm, together with details of the first line complications (infection, thrombosis or other events requiring the removal or replacement of the line) will be presented using frequencies and percentages. Information on removal of the lines as a result of the complications and the replacement of removed lines will also be provided.
These summaries will contain primarily participants randomised to the IV strategy; therefore, no statistical tests will be performed.

9.2.4 Early termination of the planned six week strategy

The frequency and percentage of participants who exited early from their allocated six week strategy for good clinical response vs other reasons (as reported on their day 42 or day 120 CRF) vs completing as planned will be presented by treatment arm and compared using chi-squared tests. If the chi-squared test indicates a difference between arms, multinomial regression will be used to estimate treatment effects of early termination for good clinical response separately from other reasons (vs completion as planned) if sufficient numbers of participants fall into this category to justify the use of a regression model.

9.2.5 Quality of life evaluated by the EQ-5D-3L questionnaire

Frequency and percentages of the number of patients within each level of the five EQ-5D-3L domains will be displayed overall and by treatment arm at baseline, 14, 42, 120 and 365 days. Descriptive statistics of the EQ-5D-3L index scores and EQ-5D-3L VAS will be presented overall and by trial arm and baseline and the relevant follow-up time points. This information will also be displayed using boxplots.

The EQ-5D-3L index score and VAS will be analysed using a quantile regression model adjusted for age, comorbidities, infecting pathogen and type of infection, as defined above. The data will be analysed separately for each follow-up time point.
As discussed in section 8.2, explanatory variables with low counts (comorbidities) and categories with low counts within explanatory variables may be combined.

9.2.6 Quality of life evaluated by the OHS and OKS (where the infection is in the hip and knee respectively)

For patients with an infection in the hip or knee, descriptive statistics will be summarised separately for the OHS and OKS overall and by treatment arm at baseline, 120 and 365 days. The data will also be displayed using boxplots.

The OHS and OKS will be analysed using separate quantile regression models adjusted for the baseline scores, age, comorbidities, infecting pathogen and type of infection, as defined above. The data will be analysed separately for each follow-up time point.
As discussed in section 8.2, explanatory variables with low counts (comorbidities) and categories with low counts within explanatory variables may be combined.

9.2.7 Adherence to oral medication

In a subset of sites (Oxford, Guys and St Thomas' Hospital Trust and Royal Free Hospital Trust) will dispense oral antibiotics in pill containers with a Medication Even Monitoring System (MEMS), whereby sensors in the pill bottle tops can detect opening and closing, and report these events with a date stamp. Results from this recording will be summarised to obtain an additional summary of adherence with the medication schedule.
Particular attention will be paid to the number of days on which all doses were missed and, within the analysis of the MEMS data, the dosing intervals. These will be analysed descriptively, using medians, interquartile ranges and ranges.

As most of the adherence data is to be completed by PO participants only, no statistical tests will be performed for these summaries.

9.2.8 Agreement between intended and received antibiotics

Agreement between the planned PO and IV antibiotics as stated prior to randomisation and actual antibiotics received will be summarised overall and by treatment arm. The frequency and percentage of participants who received and did not receive their intended treatment as their initial antibiotic regimen will be presented.

Agreement between intended and received antibiotics are categorised as follows:

Full match - received their randomised strategy and remained within the intended antibiotic group

Partial match - received their randomised strategy but deviated from the intended antibiotic group

No Match =early exit from randomised strategy

9.2.9 Antibacterial agents used for treatment

Actual initial antibiotic regimens will be summarised overall and by treatment arm. Each regimen will be classified according to the table in section 8.5.4 and summarised overall and by treatment arm.

Interruptions and changes to initial antibiotic regimen will also be tabulated overall and by treatment arm.

The number of patients continuing long-term antibiotic treatment (after 6 weekswill also be summarised overall and by treatment arm using frequencies and percentages.

Time to permanent discontinuation of all antibiotic treatment (defined as the first day where antibiotics are not taken for the next 14 days) will be compared by treatment arm using Kaplan-Meier curves.

9.2.10 Duration of primary hospital stay

Time from randomisation to discharge, and time from original admission to discharge, will be summarised overall and by treatment group using median (IQR) and compared using ranksum tests.

(Note: re-admission post-discharge is an SAE and would be presented as a secondary endpoint)

9.3 *Resource Use and Cost Data*

A separate analysis plan for the health economics analysis will be written by the trial health economist. Resource use and cost data will only be assessed for the final analysis, but not for the interim analysis.

10. Additional Analyses

10.1 *Exploratory analyses*

If the trial results do not demonstrate non-inferiority of PO, additional analyses will explore differences in the primary outcome for different levels of adherence to the oral antibiotics.

No other additional exploratory analyses are currently planned. If the trial team, in discussion with the DMC or TSC intends to perform any additional analyses, the statistical analysis plan will be updated accordingly. Any exploratory analysis that has not been pre-specified will be clearly marked as such in the final statistical report.

10.2 *Blinded analysis*

N/A – the trial statistician will not be blinded to treatment allocations while preparing and performing the statistical analysis for this trial.

10.3 *Meta-analyses*

No new meta-analysis using the trial results is planned as part of the final analysis, and the trial team are not aware of any new comparable trials in adults.

11. Safety Analysis

SAEs are collected as part of the secondary endpoints and all relevant analysis is details in section 9.

12. Appendix:

12.1 *Glossary of abbreviations*

CI	Chief Investigator
DMC	Data Monitoring Committee
ITT	Intention to Treat
IV	Intravenous antibiotics
MI	Multiple imputation
MITT	Modified Intention to treat
PO	Per Oral antibiotics
PP	Per protocol
SAP	Statistical Analysis Plan
TSC	Trial Steering Committee

12.2 *EQ-5D-3L scoring details*

The EQ-5D-3L questionnaire used in this study consists of five questions with three levels each, which are scored 1 to3, with 3 indicating the most severe problems. The five domains can be converted to a summary index using a country specific value set. Many statistical programmes include code to perform these calculations.
More detail on this questionnaire and related information can be found within the relevant scoring manual on the EuroQol Group webpage(8).

12.3 *OHS/ OKS scoring details*

The OKS and OHS consist of 12 questions each. Each item has five levels, which are scored from 0 to 4, with 4 being the best outcome. The overall score is calculated by adding up the scores for all 12 items.
If data is missing for one or two items, these values can be replaced by the mean value of all other responses. The overall score cannot be calculated if more than two items are missing.
The paper by Murray et al (2007)(9) can be referred to for additional detail.

13. Document history

Version number Issue date	Author	Significant changes from previous version
V2.0_03Dec2016	Ines Rombach	Implemented changes in line with the updated sample size calculation, and to reflect the updated non-inferiority margin (increased from 5% to 7.5%) in the primary non-inferiority analysis Updated the primary analysis to use a multiple imputation approach.

14. References

1. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal: 2013 [Accessed 29 October 2015]. Available from: https://http://www.nice.org.uk/article/pmg9/chapter/foreword.
2. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Hoboken, New Jersey: John Wiley & Sons; 2002.
3. White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. Computational statistics & data analysis. 2010 Oct 1;54(10):2267-75. PubMed PMID: 24748700. Pubmed Central PMCID: 3990447.
4. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in medicine. 2011 Feb 20;30(4):377-99. PubMed PMID: 21225900.
5. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. Clinical trials. 2012 Aug;9(4):396-407. PubMed PMID: 22752633. Pubmed Central PMCID: PMC3428470. Epub 2012/07/04. Eng.
6. Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Statistics in medicine. 2014 Oct 30;33(24):4170-85. PubMed PMID: 24845086. Pubmed Central PMCID: 4297215.
7. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharmaceutical statistics. 2013 Nov-Dec;12(6):337-47. PubMed PMID: 23292975.
8. Oemar M, Oppe M. EQ-5D-3L User Guide - Basic information on how to use the EQ-5D-3L instrument, Version 5.0 2013 [cited 2014 01 October]. Available from: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-3L_UserGuide_2013_v5.0_October_2013.pdf.
9. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and knee scores. The Journal of bone and joint surgery British volume. 2007 Aug;89(8):1010-4. PubMed PMID: 17785736.