Statistical Analysis Plan

ROADMAP: RandOmised Arthroplasty infection worlDwide Multidomain Adaptive Platform trial - analysis plan

Investigator initiated, Randomised Embedded Multifactorial Adaptive Platform (REMAP) trial, conducted across multiple hospitals in several regions of the world.

Version: 0.1

Date: 26 May 2025

Authors: Mark Jones and James Totterdell

Sponsor:	University of Newcastle, NSW, Australia
Registration (ANZCTR):	todo
HREC	todo
Chief Investigator(s):	Professor Joshua Davis Professor Laurens Manning

Table of contents

Study title:	ROADMAP: RandOmised Arthroplasty infection worlDwide Multidomain Adaptive Platform trial
Intervention:	Surgery type, backbone antibiotic duration, extended prophylaxis, antibiotic type
Study design:	Randomised Embedded Multifactorial Adaptive Platform trial
Sponsor:	University of Newcastle, NSW, Australia
Protocol:	Version 1.1 01AUG2024
Registration:	todo
HREC:	todo
Study date of first consent:	todo
Principal coordinating investigators:	Professor Joshua Davis and Professor Laurens Manning

Version history

Version	Date	Change	Reason
0.1	09/2024	First version	N/A
0.2	12/2024	Review 1	Edits based on initial comments from TS and JT

Repository status

```
## Local:    main /Users/mark/Documents/project/roadmap/src/roadmap-sap
## Remote:   main @ origin (https://github.com/maj-biostat/roadmap-sap.git)
## Head:     [af61397] 2025-05-26: Update subgroups
##
## Branches:         1
## Tags:             0
## Commits:         19
## Contributors:     1
## Stashes:          0
## Ignored files:    2
## Untracked files: 27
## Unstaged files:   2
## Staged files:     0
##
## Latest commits:
## [af61397] 2025-05-26: Update subgroups
## [3b61110] 2025-05-22: Updates to estimand based on chat with jt
## [704d6e0] 2025-05-08: Latest rendered version
## [5c89d25] 2025-05-08: Commit to push recent
## [36ffcf3] 2024-10-14: MJ review and edits for draft:
```

Preface

This statistical analysis plan (SAP) outlines the planned data processing and analysis approach for ROADMAP. The SAP is an operational document that will be updated, as necessary, prior to each analysis.

For each analysis conducted, there will be a dedicated SAP release within the github repository and the applicable release will be referenced within the analysis report.

The SAP references:

- ROADMAP Core Protocol_V1.1_01Aug2024
- ROADMAP Surgical Late Acute DSA. V1.1_01Aug2024
- ROADMAP AB Duration DSA_Part A. V1.1_01Aug2024
- ROADMAP AB Duration DSA_Part B. V1.1_01Aug2024
- ROADMAP AB Choice DSA_V1.1_01AUG2024
- ROADMAP Registry Appendix. V1.1_01Aug2024

1 Introduction

ROADMAP is a platform trial for pragmatically evaluating the effectiveness of multi-modal interventions used in the treatment of prosthetic joint infection (PJI). The primary application is to inform clinical decision making at the point of care. It is an international study run in Australia, New Zealand, Canada and the UK with partners from the academic sector. It is also designed to be a perpetual study and uses a single Master Protocol to dictate procedures and conduct.

Each set of interventions (domain) are detailed within separate Domain Specific Appendicies (DSA) to the Master Protocol. Conclusions and reports will be produced separately with some participant outcomes being used in multiple reports, thereby potentially contributing to the inference associated with several interventions. All DSAs will be evalauated under a single model specification and a single analysis plan (this document) is used cover all DSAs.

1.1 Background motivating research question(s)

ROADMAP is motivated by several high-level questions relating to the management of prosthetic joint infections, specifically:

1. For patients with *late-acute*[1] post-operative infections, is it better to keep the existing prosthetic joint and clean the infection or to replace the joint entirely?
2. What is the most effective duration of backbone antibiotics for patients receiving *one-stage revision*[2] surgery?
3. Is it better to include rifampicin (or not) in the backbone antibiotic treatment regimen?

Each of these will be evaluated in relation to the primary outcome, which is a composite outcome defined at 12-month post platform entry. Specifically, the co-occurrence of the following

[1] Page 15 of ROADMAP Core Protocol_V1.1_01Aug2024_clean.pdf says "For early post- operative infections". Think this would be clearer to refer to cohort.

[2] Page 15 of ROADMAP Core Protocol_V1.1_01Aug2024_clean.pdf says "What is the most effective duration of antibiotics after surgery?". Think this would be clearer to refer specifically to the type of surgery.

states (paraphrased from the protocol, which has a more complete definitions, see section 6.6 Trial endpoints in the Core Protocol):

1. Alive
2. No evidence of infection
3. No ongoing use of antibiotics for the index joint
4. Destination prosthesis in place

Missing elements may impact the construction of the composite primary outcome; missingness does not equate to refutation. Missingness could arise as a result of the source records being dated outside the permissible time windows, which is predominantly not earlier than 12 months post platform entry and up to 13 months post platform entry, although variations on this exist e.g. for antibiotic use.

1.2 Objectives and endpoints

The primary goal of ROADMAP trial is to examine the effect of a range of interventions in platform-eligible patients with PJI on treatment success as defined earlier, see Section 1.1. Table 1.1 provides a high-level summary of each of the domain objectives. Additional detail can be found in the protocol.

Table 1.1: ROADMAP domain-level analytic objectives

Domain	Objective	Endpoint
Surgical	Determine whether revision (understood as a sample-weighted average of the effects of one and two-stage revision) is more effective than debridement for treating prosthetic joint infections in late-silo units.	Treatment success at 12 months (alive, no evidence of infection, no antibiotics for PJI, destination prothesis in place).
Antibiotic duration	Determine whether 6 weeks of backbone antibiotic is non-inferior to 12 weeks in curing prosthetic joint infections in units receiving one-stage revision.	Treatment success at 12 months.
Extended prophylaxis	Determine whether 12 weeks of extended prophylaxis is more effective than none for treating prosthetic joint infections in units receiving two-stage revision.	Treatment success at 12 months.
Antibiotic choice	Determine whether the addition of rifampicin is more effective than none for treating prosthetic joint infections in units where one or more of the causative organisms is a Gram-positive type of interest (or infection is culture negative).	Treatment success at 12 months.

1.3 Estimands

One of the goals of the estimand framework is to create a bridge between the clinical question and the statistical analyses. In the following sections, we provide further detail to the specifications provided in the core protocol and DSAs with a focus on the treatment regimen, the population summaries and the intercurrent events.

1.3.1 Primary estimand

A primary estimand, which aims to introduce the treatment-level comparisons for all the domains, is defined in the core protocol. The DSAs introduce a domain-specific primary estimand by adding some detail on the treatment arm comparisons. For all domains, the primary estimand adopts a mixed approach to handling ICEs. The majority of ICEs are handled via a

treatment policy, which parallels the intention to treat (ITT) perspective, but the outcome definition implies a composite strategy for death[3]. In general, all ICEs are effectively considered to be part of the treatment regimen of interest and therefore receive no further consideration, Mallinckrodt et al. (2020).

1.3.1.1 Estimand B.1 (surgical intervention)

High-level overview:

The estimand considers the effect of removing the infected joint (irrespective of whether knee or hip) versus cleaning and leaving the joint in place on treatment success at 12 months (primary outcome), specifically within the cohort of patients with late-acute infection and

…where treatment in other domains are given according to the trial design dependencies and usual practice when in receipt of DAIR and when in receipt of revision (and for the type of revision adopted).

Treatment regimen:

The protocol definitions for the surgical interventions are given in the DSA for the Surgical strategy domain, which is specific to the cohort of patients with late-acute infection. Those definitions provide detail for the protocolised interventions under consideration. For example, at a minimum, debridement and implant retention (DAIR) must include an open approach, exchange of modular components, removal of the synovial lining of the joint and irrigation (see DSA Surgical Late-acute, section 7.3 for further detail).

The randomised treatment arms comprise DAIR and surgical revision where the clinician is permitted to self-select the type (one or two-stage) when a patient is assigned to the revision group. The type of revision is determined following the surgical procedure where the surgeon is required to state whether there are no plans for a second stage operation (implying a one-stage revision, see DSA for AB duration part A, section 7.2) or there is an intention to have a second stage procedure (implying two-stage revision, see DSA AB Duration part B, section

[3]While death (occurring prior, during or after the occurrence of the intervention) may ordinarily be considered an intercurrent event in this clinical setting, it has been accounted for by its inclusion in the primary outcome definition, i.e. we implicitly adopt a composite strategy for death. Similarly, secondary or rescue surgery leading to the removal of a prosthetic is accounted for by its inclusion in the primary outcome definition and again implies the composite strategy. These subtleties have to be kept in mind when interpreting and generalising the results.

6.2.1). Receipt of each of the surgical interventions and specific revision types has implications for the care revealed and admitted under other domains bar the antibiotic choice domain, although all obviously necessitate survival (a post-randomisation event) through the surgical procedure.

Receipt of DAIR precludes entry into either of the backbone antibiotic duration domain or extended prophylaxis domain and therefore, patients receiving DAIR receive non-randomised treatment (usual care) for these. However, given that DAIR is a single stage procedure, these patients have no logical progression to extended prophylaxis as is defined for two-stage revision patients[4]. DAIR patients also receive an antibiotic choice domain assignment (or non-randomised care).

Receipt of one-stage revision permits (but does not require) entry into the backbone antibiotic duration domain and precludes entry into the extended prophylaxis domain. Similar to DAIR patients, these patients have no logical progression to extended prophylaxis as is defined for two-stage revision patients. Therefore, patients receiving one-stage revision will also receive either non-randomised treatment (usual care), 12 weeks or 6 weeks backbone antibiotic duration along with an antibiotic choice domain assignment.

Receipt of two-stage revision permits (but does not require) entry into the extended prophylaxis domain and precludes entry into the backbone antibiotic duration domain. Therefore, patients receiving two-stage revision will either receive non-randomised treatment, no-extended prophylaxis or 12 weeks extended prophylaxis duration. These patients will also receive non-randomised treatment (usual care) for the backbone antibiotic duration along with an antibiotic choice domain assignment.

Intercurrent events:

Likely intercurrent events for the surgical domain and the approaches used to handle them are summarised in Table 1.2. The treatment policy strategy allows for the randomised and revealed surgical treatment in addition to a specific combination of randomised treatments and non-randomised treatments associated with other domains, and, with any occurrence of specified or unspecified ICE(s). The comparison of groups allow for lack of adherence, adding or switching, changes in background medication. The estimand is therefore not associated with a specific single treatment protocol, but rather reflects a mix of procedures that commonly occur when undertaking DAIR or revision.

[4]For the purposes of the study, extended prohylaxis is only relevant following the second stage of the revision procedure

Table 1.2: Surgery domain intercurrent events

ID	Treatment	ICE	Class	Strategy
B.1(i1)	DAIR	If prosthesis is found to be loose when the joint is opened, the surgeon would routinely override allocation and do a revision. Expected in x% of patients.	Switchover[1]	Treatment policy
B.1(i2)	Revision	If the patient becomes clinically unstable prior to or during the operation, the surgeon may abandon the attempt at revision and revert to a DAIR. Expected in x% of patients.	Switchover	Treatment policy
B.1(i3)	Revision	If one or more components of the prosthesis are too difficult to remove, the surgeon may abandon the attempt at revision and revert to a DAIR. This is more likely with a non-cemented hip than with a cemented hip or a knee. Expected in x% of patients.	Switchover	Treatment policy

[1]Switchover will not generally result in the outcome variable being missing (assuming the patient followup continues to completion) but does affect the interpretation of the result due to dilution of the treatment effect. For an (admittedly contrived) example, in a two-arm trial, if the majority of units assigned to the first treatment switched over to the second treatment and none of the units assigned to the second treatment switched then an ITT/treatment policy approach would be comparing groups of units that effectively received the same treatment.

Summary measure:

The population-level summary measure is the absolute risk difference with respect to treatment success comparing revision (defined as both one-stage and two-stage procedures) and DAIR in conjunction with the set of treatments in other domains as determined by the design and usual care by the relevant surgical procedure assigned.

Main estimator:

A Bayesian multivariable logistic regression model is used to characterise unit responses for the primary outcome across all domains, see Section 2.6. The approach is intended to (1) accommodate clinician-choice with regards to revision type and (2) allow for the dependency between specific revision types and entry into the antibiotic duration and extended prophylaxis domains with differing baseline log-odds.

The population-level summary is obtained via g-computation, which provides a counterfactual comparison across the relevant cohort. We estimate the unit-level probability of treatment success under each surgical intervention type and average over the sample. A Bayesian bootstrap procedure accounts for uncertainty in the covariate distribution. The counterfactual assignments across the domains are obtained by reference to the baseline randomisation. The overall

revision effect is the risk difference between a weighted combination of the one and two-stage revision risk and DAIR where the weights arise from the sample distribution for the surgical revision preference under assignment to revision, obtained at baseline and restricted to the late acute silo. Any varying effects will be fixed at zero rather than marginalised.

Data useful for estimand:

The ability to compute the estimand assumes we have access to each patients' unconditional random allocation, across all domains, which is made at baseline (i.e. at platform entry). Crucially, patients' allocated to DAIR in the surgical domain will be randomised within the backbone antibiotic duration and extended prophylaxis duration domains, but these allocations would likely never be revealed.

Surgical preference for revision type is required for all participants, at baseline. Specifically, we require an answer to "Assuming the patient receives revision, what type of revision procedure is preferenced at this time?"

Any ICE should be detailed with respect to reason, timing and implications.

Patients should be followed up after the occurrence of an ICE to obtain detail on the final status of the primary outcome.

Missing data:

Missing data might arise due to limitations in identifying clinical records, formal requests from patients to completely withdraw from the study (including use of data), loss to follow up due to relocation, human error etc. Additionally, components of the composite outcome may not be available in the clinical records (e.g. ongoing use of antibiotics for the specific reason of managing prosthetic infection might not be adequately recorded). Missingness will be assumed to be low and at random across all variables (and/or variable components) such that imputation could, in principle, be undertaken by conditioning on observed patient characteristics. Under the assumption that the missingness is not simultaneously dependent on the outcome and exposure, multiple imputation is not technically required to obtain unbiased estimates of the treatment effect. Additionally, a complete-case (likelihood-based) analysis will yield unbiased estimates of the treatment effects. As such, multiple imputation will not be adopted for the primary outcome variable. Where it occurs, missingness in covariates will be addressed by deduction (where possible) or fixed imputation scheme (if reasonable) the details of which are to be disclosed in analysis reports and publications.

Sensitivity:

Tipping point sensitivity analysis and multiple imputation schemes may be considered and implemented at the discretion of the analyst, the details of which being disclosed in the reporting.

1.3.1.2 Estimand D.1 (antibiotic duration intervention)

High-level overview:

The estimand considers the effect of the backbone antibiotic duration on treatment success at 12 months, specifically whether 6 weeks (a short duration) of backbone antibiotic is non-inferior to 12 weeks (a long duration) in patients surviving the first stage of revision with no plans for a secondary stage (and)

…where treatment in other domains are given according to the trial design and usual practice under entry into the backbone antibiotic duration domain.

The domain effectively represents a sub-study based on patients that received survive a specific type of revision procedure.

Treatment regimen:

The protocol definitions of the interventions are provided in the DSA for the antibiotic duration domain following single stage revision.

Entry into the domain is determined based on receipt and confirmation of one-stage revision, which occurs on the conclusion of the surgery by a surgeon indicating that no secondary stage is intended (see DSA for AB duration part A, section 7.2).

Eligibility for the backbone antibiotic duration domain precludes entry into the extended prophylaxis domain, as there is no progression to a secondary stage procedure (per the form used under two-stage revision) for this cohort of patients. Patients entering into the backbone antibiotic duration domain also receive an antibiotic choice domain assignment (or non-randomised care).

Intercurrent events:

Potential intercurrent events and the approaches used to handle them are summarised in Table 1.3 where the treatment policy ICE strategy aims to establish the effect of treatment assignment.

Table 1.3: Antibiotic duration domain intercurrent events

ID	Treatment	ICE	Class	Strategy
D.1(i1)	6 wk	If patient has slow improvement, recrudescence of infection or need to return to theatre then clinicians may choose to prolong antibiotic therapy. Expected in x% of patients.	Extension to therapy[1]	Treatment policy
D.1(i2)	12 wk	Adverse effects of antibiotics or patient discharge/relocation may lead to early termination of therapy. Expected in x% of patients.	Discontinuation[2]	Treatment policy

[1]Extended therapy in the 6 week group would make the responses more similar to the 12 week group leading to increase likelihood of non-inferiority decision, i.e. conclusions may anti-conservative.

[2]Discontinuation in the 12 week group would make the responses more similar to the 6 week group.

The treatment regimen under evaluation is thus the randomised and revealed antibiotic duration treatment in conjunction with any occurrence of specified or unspecified ICE(s) in combination with randomised and non-randomised treatments associated with other domains. That is, interest is in the comparison of groups without allowances for lack of adherence, adding or switching, changes in background medication, and so on.

Summary measure:

The population-level summary is an absolute risk difference with respect to treatment success comparing 6 weeks backbone antibiotic and 12 weeks in patients receiving one-stage revision in conjunction with the treatments in other domains as determined by the design dependencies and usual care.

Main estimator:

Similar to the estimator for the surgical domain, the population-level summary is obtained via g-computation, which provides a counterfactual comparison across the relevant cohort, here restricted to patients in recipt of one-stage revision under a landmark analysis. Again, we estimate the unit-level probability of treatment success under each randomised intervention and average over the sample. A Bayesian bootstrap procedure to account for uncertainty in the covariate distribution.

Data useful for estimand:

Per the considerations and definitions for B.1.

Missing data:

Per the considerations and definitions for B.1.

Sensitivity:

Per the considerations and definitions for B.1.

1.3.1.3 Estimand E.1 (extended prophylaxis intervention)

High-level overview:

The estimand considers the effect of the extended prophylaxis on treatment success at 12 months, specifically whether 12 weeks of extended prophylaxis is more effective than none following the second stage within a two-stage revision procedure and

…where treatment in other domains are given according to the trial design dependencies and usual practice when in receipt of two-stage revision.

Treatment regimen:

The protocol definitions of the interventions are provided in the DSA for the Antibiotic duration domain (extended prophylaxis) following a two-stage revision (see DSA AB Duration part B, section 6.3).

Entry into the domain for randomised treatment is determined based on the adoption of two-stage revision and survival through the second stage of surgery. Determination of two-stage revision is made following the first stage operation (see DSA AB Duration part B, section 6.2.1) and the reveal to extended prophylaxis randomised treatment is made following the second stage operation. This has obvious implications on the necessity to survive through the intervening period. In other words, if a patient is intended to have a two-stage revision, but does not survive to and through the second procedure, then they are not revealed to their extended prophylaxis assignment and do not contribute to the randomised comparison.

While receipt of two-stage revision precludes entry into the backbone antibiotic domain, patients in this domain also receive an antibiotic choice domain assignment (or non-randomised care).

Intercurrent events:

Potential intercurrent events and the approaches used to handle them are summarised in Table 1.4 where the treatment policy ICE strategy aims to establish the effect of treatment assignment.

Table 1.4: Extended prophylaxis domain intercurrent events

ID	Treatment	ICE	Class	Strategy
E.1(i1)	None	If the patient has slow improvement, recrudescence of infection or need to return to theatre then clinicians may choose to prolong extended prophylaxis therapy. Expected in x% of patients.	Extension to therapy[1]	Treatment policy
E.1(i2)	None	If the patient shows late positive culture, they may become ineligible post-reveal and would likely require an additional 6-12 weeks of backbone antibiotic. Expected in x% of patients.	Extension to therapy	Treatment policy
E.1(i3)	12 wk	Adverse effects of antibiotics or patient discharge/relocation may lead to early termination of therapy. Expected in x% of patients.	Discontinuation[2]	Treatment policy

[1]Extended therapy in the no extended prophylaxis group would make the group more similar to the 12 week group leading to increase likelihood of non-inferiority decision, i.e. conclusions may anti-conservative.
[2]Discontinuation in the 12 week group would make the responses more similar to the no extended prophylaxis group.

The treatment regimen under evaluation is thus the randomised and revealed extended prophylaxis treatment in conjunction with any occurrence of specified or unspecified ICE(s) in combination with randomised and non-randomised treatments associated with other domains. That is, interest is in the comparison of groups without allowances for lack of adherence, adding or switching, changes in background medication, and so on.

Summary measure:

The population-level summary is an absolute risk difference with respect to treatment success comparing 12 weeks extended prophylaxis and none in patients receiving two-stage revision in conjunction with the treatments in other domains as determined by the design dependencies and usual care.

Main estimator:

The population-level summary is obtained via g-computation, which provides a counterfactual comparison across the relevant cohort, here restricted to patients in receipt of two-stage revision under a landmark analysis. We estimate the unit-level probability of treatment success under each randomised intervention in the extended prophylaxis domain and average over the sample A Bayesian bootstrap procedure to account for uncertainty in the covariate distribution.

Data useful for estimand:

Per the considerations and definitions for B.1.

Missing data:

Per the considerations and definitions for B.1.

Sensitivity:

Per the considerations and definitions for B.1.

1.3.1.4 Estimand C.1 (antibiotic choice intervention)

High-level overview:

The estimand considers the effect of the antibiotic choice on treatment success at 12 months, specifically whether the use of rifampicin is more effective than no rifampicin following the initial surgical procedure (irrespective of type) and

…where treatment in other domains are given according to the trial design dependencies and usual practice specific to the surgical procedure received.

Treatment regimen:

The protocol definitions of the interventions are provided in the DSA for the Antibiotic choice domain (see DSA AB Choice, section 7.3).

Entry into the domain for randomised treatment is again predicated on survival through the initial surgical procedure, i.e. after a DAIR, one-stage or between the first and second stage of a two-stage revision.

Given that the reveal process for antibiotic choice will occur after the revision type (one or two-stage) has been nominated by the surgeon, it cannot influence the surgeon's decision.

Intercurrent events:

Potential intercurrent events and the approaches used to handle them are summarised in Table 1.5 where the treatment policy ICE strategy aims to establish the effect of initially randomised treatment.

Table 1.5: Choice domain intercurrent events

ID	Treatment	ICE	Class	Strategy
C.1(i1)	None	If the patient has slow improvement (e.g. 1-4 weeks post operatively), clinicians may add rifampicin despite assignment. Expected in x% of patients.	Switchover	Treatment policy
C.1(i2)	Rifampicin	If the site pharmacy runs out of or does not have any rifampicin in stock, this may lead to early termination or non-receipt of therapy. Expected in x% of patients.	Discontinuation or non-receipt[1]	Treatment policy
C.1(i3)	Rifampicin	Adverse effects of rifampicin (intractable nausea, severe hepatitis) may lead to early termination of therapy. Expected in x% of patients.	Discontinuation[1]	Treatment policy

[1]Both discontinuation/non-receipt of rifampicin or addition of rifampicin to the the control group make groups more similar than they otherwise would be.

The treatment regimen under evaluation is thus the randomised and revealed antibiotic choice treatment in conjunction with any occurrence of specified or unspecified ICE(s) in combination with randomised and non-randomised treatments associated with other domains. That is, interest is in the comparison of groups without allowances for lack of adherence, adding or switching, changes in background medication, and so on.

Summary measure:

The population-level summary is an absolute risk difference with respect to treatment success comparing rifampicin and no rifampicin following the initial procedure in conjunction with the treatments in other domains as determined by the design dependencies and usual care.

Main estimator:

The population-level summary is obtained via g-computation, which provides a counterfactual comparison across the relevant cohort under a landmark analysis. We estimate the unit-level probability of treatment success under each randomised intervention for the antibiotic choice domain and average over the sample. A Bayesian bootstrap procedure to account for uncertainty in the covariate distribution.

Data useful for estimand:

Per the considerations and definitions for B.1.

Missing data:

Per the considerations and definitions for B.1.

Sensitivity:

Per the considerations and definitions for B.1.

1.3.2 Supportive estimands

Supportive estimands are provided to address the needs of other stakeholders giving alternative perspectives and interpretations of intervention effects.

Given the number of estimands defined in the DSAs, only those where ICEs are not handled under the treatment policy strategy are discussed.

1.3.2.1 Estimand B.2 (surgical intervention)

High-level overview:

Per the primary estimand for the surgical domain, this supportive estimand considers the effect of removing the infected joint versus cleaning and leaving the joint in place on treatment success at 12 months. However, the treatment regimen under consideration is tightened to a sub-population that adhered to the assigned treatment.

Principal stratum via mixture model but very challenging given the dimensionality of the problem and the limited specification.

Treatment regimen:

Intercurrent events:

ICEs are per the primary estimand for the surgical domain.

Summary measure:

todo

Main estimator:

The population-level summary is obtained via g-computation, which provides a counterfactual comparison across the relevant cohort under a landmark analysis. We estimate the unit-level probability of treatment success under each randomised intervention for the antibiotic choice domain and average over the sample. A Bayesian bootstrap procedure to account for uncertainty in the covariate distribution.

Data useful for estimand:

It is important that units continue to be followed up after the occurrence of an ICE to obtain detail on the final status of the primary outcome.

Record which of the patients, assigned to revision, had loose joints such that had they been assigned to DAIR, then switchover would likely have occurred.

Record which of the patients, assigned to DAIR, had indications of prosthesis that would likely be difficult to remove, such that had they been assigned to revision, then switchover would likely have occurred.

Record if the patient became unstable or neared instability at any point during the procedure for both DAIR and revision.

1.3.2.2 Estimand D.2 (antibiotic duration)

High-level overview:

This supportive estimand considers the effect of the duration of backbone antibiotics on treatment success at 12 months, specifically whether 6 weeks (a short duration) of backbone antibiotic is non-inferior to 12 weeks (a long duration) in patients receiving one-stage revision. However, the treatment regimen under consideration is tightened to a sub-population that adhered to the assigned treatment.

todo - finalise approach

Treatment regimen:

Intercurrent events:

Summary measure:

Main estimator:

Data useful for estimand:

Missing data:

Sensitivity:

1.3.2.3 Estimand E.2 (extended prohylaxis)

todo - pp finalise approach

Treatment regimen:

Intercurrent events:

Summary measure:

Main estimator:

Data useful for estimand:

Missing data:

Sensitivity:

1.3.2.4 Estimand C.2 (antibiotic choice)

todo - pp finalise approach

Treatment regimen:

Intercurrent events:

Summary measure:

Main estimator:

Data useful for estimand:

Missing data:

Sensitivity:

1.4 Study design

ROADMAP is an investigator-initiated, phase IV, open label, multicentre, pragmatic, randomised embedded multifactorial adaptive platform (REMAP) formulated to investigate the effectiveness of multiple study interventions simultaneously in cohorts of patients with confirmed or likely prosthetic joint infection in a large joint (hip, knee) with no age restriction. ROADMAP also includes the development of a registry, although that will not be discussed here.

Initial treatment modality groups (domains) examine surgery type, antibiotic duration, extended prophylaxis and antibiotic choice. New interventions are permitted to enter into existing domains and new domains are also permitted, both subject to steering committee and ethics review.

ROADMAP will be conducted sequentially (as cohorts of 500 patients reach their primary endpoint). Decision rules evaluated on parameter estimates of interest, which drive domain-level stopping rules and platform conclusions. Early stopping is permitted under pre-specified conditions, specifically for superiority, non-inferiority and futility as applicable to the given domain.

Bayesian methods were selected for their flexibility, ease of uncertainty quantification and their capacity for incorparating adaptive elements, as well as regularisation of parameter estimates and relatively straight forward interpretation.

1.4.1 Randomisation

Units will be unconditionally randomised (under fixed complete randomisation, i.e. non-adaptive and without restriction through blocking or stratification or other constraints) to one arm within every domain (Rosenberger and Lachin, 2016). The method was selected for its operational simplicity.

Domain specific assignments are revealed based on the design dependencies. For example, only units in receipt of one-stage revision are revealed to randomised assignment for antibiotic duration and this also precludes entry into extended prophylaxis. Similarly, only units in receipt of two-stage revision can be revealed to randomised assignment for extended prophylaxis. The reveal process actuates the randomised treatment and, it is not until this occurs, that the unit will contribute to the inference for the randomised comparison.

As the study progresses, decisions may occur at a domain level that prevent subsequent reveal for applicable treatment arms. However, patients that have entered the platform but have not been revealed at the time of a stopping decision will, in general[5], retain their original assignment.

1.4.2 Sample size

While the study is intended to be perpetual, the initial trial funding and infrastructure has sufficient resources to enrol up to 2,500 participants into the platform. The sample size is therefore constrained by the available resources and the desired trial structure. The actual sample size may be lower than 2,500 if stopping rules are invoked. Further details of the expected sample size under various scenarios can be found in the simulation report.

1.4.3 Data management

An overview of data management procedures is provided in the Master Protocol (section 8.12) with some further detail in the Registry appendix (section 7.6).

The data storage approach will be decomposed into redcap components and (out-sourced proprietary) platform components developed by Spiral Software[6] with source data (obtained from various medical records or direct report) entered by site personnel. Spiral are also responsible for the implementation of the randomisation processes.

[5] Are there exceptions where this would not be the case?
[6] https://spiral.co.nz/

2 Statistical methods

2.1 Analysis sets

2.1.1 Intention to treat

The intent to treat (ITT) principle address what participants and what data to include on each person entering the study. A strict interpretation of ITT demands collection and analysis of all randomised units, but in practice minor deviations from this are routinely accepted.

Estimands B.1 through C.1 align with the ITT perspective through their use of the treatment policy strategy. The analysis population used for these estimands will be referred to as the full analysis set and comprise all units that were randomised and revealed to at least one of the domain interventions and have passed the primary endpoint of 12-months. Per the treatment policy strategy, all randomised patients will be included and analysed according to the regimen to which they were initially allocated irrespective of any deviations from this regimen or any other protocol deviations.

Participants that have reached follow up, but for whom information has not yet been gathered will be treated as missing and excluded from analyses until the data has been entered.

2.1.2 Per-protocol

For supportive estimands, we define a per-protocol analysis comprising participants that completed protocolised progression through the study without deviation. That is, both participant and clinical team adheres with the predefined procedures, criteria and/or timelines detailed in the protocol that span consent, eligibility, adherence to randomised (and non-randomised) interventions, data collection and safety. The minimal requirements for determining that patients met protocolised treatment criteria (i.e. the accepted delivery of intervention) are provided in Table 2.1. Further detail is available in the relevant DSAs.

Clinical research staff will review patient records in order to identify the per-protocol population. Ideally, the review procedures will be pre-specified and undertaken by an independent subject matter expert.

Table 2.1: Minimal requirements for per protocol population

Domain	Assignment	Requirement
Surgical	DAIR	Part or all of the index prosthesis was retained, and an open arthrotomy was performed, including synovectomy, lavage and exchange of modular components (if present), between platform entry and day 90
	Revision	The index prosthesis was completely removed, with no residual prosthetic components, and either a new prosthesis or a temporary spacer was placed at the first stage operation, between platform entry and day 90
Antibiotic duration	6 weeks	At least 5 weeks but no more than 7 weeks of antibiotic therapy has been completed between the date of the one-stage revision and 16 weeks later
	12 weeks	At least 11 weeks and no more than 13 weeks of antibiotic therapy has been completed between the date of the one-stage revision and 16 weeks later
Extended prophylaxis[1]	None	Less than 14 days of antibiotics were received for the index joint between the reimplantation operation and platform day 90
	12 weeks	10-14 weeks of antibiotics were received for the index joint between the reimplantation operation and platform day 90
Antibiotic choice	Rifampicin	At least 1 dose of rifampicin was received on each of at least 7 days between confirmation of domain eligibility and the end of platform day 28.
	None	Less than 3 doses of rifampicin were received (i.e. zero, one or two) between confirmation of domain eligibility and the end of platform day 90.

[1]Antibiotics [if the patient is still on them] are ceased within 24 hours of confirmation of allocation reveal – which will be 4-10 days post the reimplantation stage

2.2 Subgroups

Stratification of data to subgroup populations enable the exploration of effect heterogeneity, but come at the increased risk of instability, bias and false positives.

The subgroup populations have been identified in the DSAs (DSA Surgical Late-acute section 9.6, DSA for AB duration part A section 9.6, DSA AB Duration part B section 8.6, DSA AB Choice section 9.6) but descriptions are presented here in abbreviated form for convenience, see Table 2.2. The analysis approach for subgroups is detailed in Section 2.8.

Table 2.2: Subgroup populations

Domain	Subgroup
Surgical	Site of infection by joint (hip/knee) in the index prosthesis (the infected prosthesis that was present at time of platform entry and for which the patient met eligibility criteria)
	Duration of symptoms at domain entry (categorised as duration <= 7 days, 7 days < duration <= 14 days, 14 days < duration <= 21 days)
	At least one causative organism is S. aureus versus not
	Serum C-reactive protein (CRP) at platform entry <100 versus >=100
	Time from implantation of the index prosthesis to domain entry in days
	One stage versus two stage revision (in those who are allocated to revision surgery)
Antibiotic duration	Silo membership (early, late-acute or chronic)
	At least one causative organism is known at the time of domain eligibility assessment to be Staphylococcus aureus versus not
	Revision procedure has all elements of an 'ideal' procedure vs. not
Extended prophylaxis	Silo membership
	At least one causative organism is known at the time of domain eligibility assessment to be Staphylococcus aureus versus none
	Duration between first-stage and reimplantation procedure
	Duration of antibiotic treatment between first-stage and reimplantation procedure
Antibiotic choice	Type of surgery (DAIR, one-stage, two-stage)
	At least one causative organism is a Staphylococcus (any species) versus none

For the surgical domain, one versus two stage revision effects are already addressed within the primary analysis model.

2.3 Descriptive summaries

A CONSORT diagram will will be provided aiming to detail patient progression including:

- participants screened
- participants eligible by domain (giving reasons for ineligibility)
- participants consented (a single consent process covers the platform and all relevant domains, see Core protocol, section 8.4)
- participants entering into randomised treatment (revealed) by domain and intervention
- participants withdrawing from study
- participants reaching 12-month follow up

Recruitment numbers will be reported by region, site, silo. Detail on intervention availability by domain will be presented by site. Number of protocol deviations and intercurrent events will be summarised by domain and intervention.

Baseline characteristics will be provided by silo, domain and intervention including:

- age
- sex
- ethnicity
- number of comorbidities
- ?

Concomitant medications will be reported, if available.

2.4 Sequential analyses

Interim analyses will be run over the life of the trial to evaluate pre-specified decision criteria. Analyses will start once 500 participants have reached 12 months follow up and every subsequent 500 participants thereafter. Prior to each analysis, a SAP release will be created within the github repository and the release will be referenced within the analysis report.

With the exception of the clinicial team involved in administration of the treatments and patients, only the analytical and data groups, that are responsible for providing analysis results to the DSMC, will have access to individual-level treatment group assignments.

The interim analysis will focus solely on the primary outcome and will follow the approach detailed in Section 2.6. In the event of a futility trigger being met, a subgroup analysis will be run which will enable the DSMC to explore the decision of whether heterogeneity in any subgroups is sufficient to warrant continuing the study for particular groups.

Subgroup analyses may be run at each interim analysis if futility rules have being triggered, see Section 2.8.

2.5 Analysis approach

Analyses will be conducted within a Bayesian framework with a focus on the estimation of estimands, see Section 1.3. Parameter estimates will be computed via Markov chain Monte Carlo (MCMC) using Hamiltonian Monte Carlo (HMC). Posterior summaries will be reported as posterior means and medians, with 95% credible intervals and posterior standard deviations. Convergence will be assessed visually and with reference to appropriate statistics. Model fit will be considered with reference to posterior predictive checks.

2.6 Primary analysis

The primary analysis model for ROADMAP will adjust for silo, joint, preference for revision type, treatment regimen, time period of recruitment, region/site and baseline characteristics, adopting the following form (with constraints imposed for identifiability):

$$
\begin{aligned}
Y &\sim \text{Bernoulli}(p) \\
\text{logit}(p) &= \mu + \lambda_s + \rho_j + \phi_q + \\
&\quad \beta_{d_1,s} + \beta_{d_2} + \beta_{d_3} + \beta_{d_4} + \\
&\quad \tau_t + \psi_r + \zeta_{z(r)} + \delta
\end{aligned}
$$

where Y is a binary variable representing unit level treatment success with probability p and the linear predictor terms are as follows:

- μ a reference level log-odds of a successful outcome
- λ_s shift attributable to membership in silo s
- ρ_j shift attributable site of infection j

- ϕ_q shift attributable to surgeon preference for one/two stage, q, assuming unit randomised to revision
- $\beta_{d_1,s}$ shift attributable to silo-specific surgical intervention
- β_{d_2} shift attributable to backbone antibiotic duration with reference level corresponding to non-randomised /non-applicable unit level treatment
- β_{d_3} shift attributable to extended prophylaxis duration with reference level corresponding to non-randomised /non-applicable unit level treatment
- β_{d_4} shift attributable to antibiotic choice with reference level corresponding to non-randomised /non-applicable unit level treatment
- τ_t shift attributable to randomisation period t
- ψ_r shift attributable to region r
- $\zeta_{z(r)}$ shift attributable to site z nested within region r
- δ shift attributable to a sum of unit specific parameters that account for heterogeneity in baseline characteristics

Baseline characteristics include:

- presence of immune compromise
- presence of more than one comorbidity (diabetes, rheumatoid arthritis, obesity, chronic kidney disease)
- smoking
- age
- frailty
- ...

Continuous measures may be modelled via a penalised smoothing to accommodate non-linearity.

Parameters estimates will be reported as point and interval summaries of the posterior. The treatment effect estimands will be calculated from the posterior using g-computation, see Section 3.

2.6.1 Priors

We will use weakly informative priors that aim to constrain the parameter estimates to within plausible ranges and are consistent with the belief that extreme treatment effects are unlikely.

The prior for μ will be set to

$$\mu \sim \text{Logistic}(0.7, 0.7) \quad \text{location/scale}$$

Converted to the probability scale, this gives a prior median around 0.7 and puts 90% of the prior density between 0.2 and 0.9.

The priors for all main effects will be set to Normal$(0, 1)$ which are centred on zero with 90% of its mass between ± 3 on the log-odds scale.

Noting that some domains may vary their active set of interventions over time, a first-order random walk may be used to model temporal variation in the background response. The random walk prior has the following structure

$$\begin{aligned} \tau_1 &= 0 \\ \tau_i &= \text{Normal}(\tau_{i-1}, \sigma_\tau) \quad \forall\, i > 1 \\ \sigma_\tau &\sim \text{Exponential}(1) \end{aligned}$$

with indexes aligned with analyses and the τ_1 term representing the current quarter (i.e. 3 monthly intervals). The time adjustment will only be included in the model if treatment is varied over time.

The prior for region will be set as per the main effects detailed above with the first region fixed to zero. Site priors will be nested within region and set as

$$\begin{aligned} \zeta_{z(r)} &\sim \text{Normal}(0, \sigma_\zeta) \\ \sigma_\zeta &\sim \text{Exponential}(1) \end{aligned}$$

The primary model results will be assessed using posterior predictive checks. If issues with the pre-specified models arise, the necessary modifications will be reported including justification.

2.7 Sensitivity analysis (applicable to primary)

Sensitivity analyses of the results to the pre-specified priors will be run at each analysis and variations to the pre-specified models may be explored. Additional models (either simpler or more complex) may be investigated as part of checks of sensitivity, stability, and model fit.

2.8 Subgroup analyses

Pre-specified subgroup analyses will be restricted to the primary model with additional post-hoc exploratory subgroup analyses being discretionary. The general approach will be to use the complete data, incorporating first-order interactions via hierarchical modelling for each subgroup considered.

Analyses will be run at the time of final reporting for each domain, but also at interim analyses for the relevant domain-level subgroups, if a futility decision is triggered. The latter analysis motivated by a prior belief of clinically relevant treatment effect heterogeneity and a desire to mitigate the possibility of terminating entry into a domain for subgroups when the possibility of a positive outcome remained.

To address subgroups, the primary analysis model is revised such that the parameters in the linear predictor are split for all relevant groups.

$$
\begin{aligned}
Y &\sim \text{Bernoulli}(p) \\
\text{logit}(p) &= \mu + \lambda_s + \rho_j + \phi_q + \\
&\quad \beta_{d_1,s} + \beta_{d_2} + \beta_{d_3} + \beta_{d_4} + \\
&\quad \tau_t + \psi_r + \zeta_{z(r)} + \delta
\end{aligned}
$$

where the terms are per the original specification with the treatment effects redefined to incorporate heterogeneity associated with subgroup membership:

$$
\beta_{d_1,s} = \beta'_{d_1,s} + \gamma_{d_1,j} + \gamma_{d_1,a} + \gamma_{d_1,c} + \gamma_{d_1,r}
$$

where

- $\beta'_{d_1,s}$ fixed silo-specific surgical intervention effect in the reference subgroup

- $\gamma_{d_1,j}$ shift in the applicable surgical intervention associated with site of infection
- $\gamma_{d_1,a}$ shift in the applicable surgical intervention associated with duration of symptoms at entry
- $\gamma_{d_1,c}$ shift in the applicable surgical intervention associated with presence of causative organism
- $\gamma_{d_1,r}$ shift in the applicable surgical intervention associated with CRP at baseline

which, for simplicity, is assuming that the subgroup deviations are common to all silos. The priors are set to:

$$\begin{aligned}\beta'_{d_1,s} &\sim \text{Normal}(0, 1)\\ \gamma_{d_1,\cdot} &\sim \text{Normal}(0, \omega_{d_1})\\ \omega_{d_1} &\sim \text{HalfNormal}(1)\end{aligned}$$

Similarly, for the backbone antibiotic duration domain:

$$\beta_{d_2} = \beta'_{d_2} + \gamma_{d_2,s} + \gamma_{d_2,c}$$

- β'_{d_2} fixed backbone antibiotic duration intervention effect in the reference subgroup
- $\gamma_{d_2,s}$ shift in the intervention effect associated with silo membership
- $\gamma_{d_2,c}$ shift in the intervention effect associated with presence of causative organism

The priors are set to:

$$\begin{aligned}\beta'_{d_2,s} &\sim \text{Normal}(0, 1)\\ \gamma_{d_2,\cdot} &\sim \text{Normal}(0, \omega_{d_2})\\ \omega_{d_2} &\sim \text{HalfNormal}(1)\end{aligned}$$

For the extended prophylaxis domain:

$$\beta_{d_3} = \beta'_{d_3} + \gamma_{d_3,s} + \gamma_{d_3,c}$$

- β'_{d_3} fixed extended prophylaxis intervention effect in the reference subgroup
- $\gamma_{d_3,s}$ shift in the intervention effect associated with silo membership

The priors are set to:

$$\begin{aligned}\beta'_{d_3,s} &\sim \text{Normal}(0, 1)\\ \gamma_{d_3,\cdot} &\sim \text{Normal}(0, \omega_{d_3})\\ \omega_{d_3} &\sim \text{HalfNormal}(1)\end{aligned}$$

For the antibiotic choice domain:

$$\beta_{d_4} = \beta'_{d_4} + \gamma_{d_4,d_1} + \gamma_{d_3,c}$$

- β'_{d_4} fixed antibiotic choice intervention effect in the reference subgroup
- γ_{d_4,d_1} shift in the intervention effect associated with type of surgery
- $\gamma_{d_4,c}$ shift in the intervention effect associated with presence of causative organism

The priors are set to:

$$\begin{aligned}\beta'_{d_4,s} &\sim \text{Normal}(0, 1)\\ \gamma_{d_4,\cdot} &\sim \text{Normal}(0, \omega_{d_4})\\ \omega_{d_4} &\sim \text{HalfNormal}(1)\end{aligned}$$

that is, with shared variance components for the domain level subgroup terms.

2.9 Supportive (domain agnostic) analyses

The following sections detail analyses applying to all domains (hence *domain agnostic*) - interest is in treatment effects associated with all domain level interventions.

2.9.1 Desirability of outcome ranking

A desirability of outcome ranking (DOOR) analysis involves unit level comparisons between all trial participants across the treatment arms. Each patient receives a single rank to characterise their overall state at 12-months after platform entry with the current DOOR criteria provided in Table 2.3.

Table 2.3: Ranking criteria for desirability of outcome for PJI

Rank	Alive	Joint Function	Treatment Success[1]	Qol
1	Yes	Good[2]	Yes	Tiebreaker based on EQ5D5L
2	Yes	Good	No	Tiebreaker based on EQ5D5L
3	Yes	Poor	Yes	Tiebreaker based on EQ5D5L
4	Yes	Poor	No	Tiebreaker based on EQ5D5L
5	No	-	-	-

[1]Treatment Success relates to primary outcome definition.

[2] *'Good'* joint function is based on thresholds related to patient reported success. A successful outcome at 12-months will be defined for knee PJI with an Oxford Knee Score (OKS) at 12 months of >36 or an improvement (delta) from baseline of >9 and for hip PJI as a Oxford Hip Score (OHS) of >38 or an improvement of >12 (35).

The DOOR analysis approach targets what has been called a *DOOR probability* but is more broadly referred to as a Probabilistic Index (PI), see De Schryver (2019) and is equivalent to the Mann-Whitney-U statistic. To calculate the PI we will enumerate all pairwise comparisons between all patients (win, loss, tie) and then the PI is given by

$$PI = \frac{(n_{win} + 0.5n_{tie})}{n_e n_c}$$

where n_{win} and n_{tie} are the number of instances where the experimental units do *better* than the control patients or have equivalent ranking respectively for all pairs and n_e and n_c are the number of patients in each of the two arms being compared. We will produce confidence intervals via a bootstrap procedure or other methods.

We note that the PI refers to the probability that the outcome of a randomly selected subject in one group exceeds the outcome of another randomly selected subject in another group (plus half of the probability of a tied DOOR), see Evans and Follmann (2016). In principle, it targets:

$$\Pr(Y_e > Y_c)$$

where Y_e is a random variable describing the outcome under an experimental treatment and Y_c is a corresponding outcome under the control.

Separate DOOR analyses will be run on strata based on the assessment and ranking of the relevant outcomes at 12-months. That is, in the DOOR analysis, we are comparing a subset of units over the interventions specific to a given domain. Specifically:

- For the surgical domain, the PI will be computed for only the units within the late-acute silo revealed to randomised treatment for this domain
- For the antibiotic duration domain, the PI will be computed for only the units that received one-stage revision and revealed to randomised treatment for this domain
- For the extended prophylaxis domain, the PI will be computed for only the units that received two-stage revision and revealed to randomised treatment for this domain
- For the antibiotic choice domain, the PI will be computed for all units revealed to randomised treatment for this domain

2.9.2 Patient-reported joint function

The Oxford Hip Score (OHS) is a joint-specific, patient-reported outcome measure that has been designed to assess disability in patients undergoing joint replacement. The score is computed based on the responses to a 12-item survey. Under the 2007 specification, responses range from 0 to 4 for each question and the total score has a maximum (best) value of 48 with 40-48 indicating satisfactory joint function and 3-5 being a suggested clinically important difference.

The OHS will be analysed using cumulative logistic regression based on the data obtained at 12-months after platform entry. Letting $Y \in \{1, 2, \dots K\}$ denote the outcome (with $K = 49$ here, accounting for death as the lowest level) for unit i, the proportional odds model can be considered with reference to categorising some latent continuous variable Y^*. Ordered cut-points, $c \in \mathbb{R}^{K-1}$ are defined such that $c_k < c_{k+1}$ with

$$\text{logit}(\Pr(Y \leq c)) = c_k - x^\top \beta$$

and where the linear predictor (generically stated in the above) would have terms and priors analogous to those used in the primary analysis. The priors for the intercept terms would be based on a weakly informative dirichlet prior.

2.9.3 Patient-reported quality of life (EQ5D5L)

The the EuroQOL 5 dimension 5 levels (EQ-5D-5L) instrument is a preference-based QoL instrument comprised of five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Respondents are asked to choose the most appropriate option from five alternatives (none, slight, moderate, severe, or extreme problems). In addition, respondents are asked to indicate their present health state on a visual analogue scale (EQ VAS) ranging from the worst imaginable health state ("0") to the best imaginable health state ("100").

At this time, we assume that the QoL at 12-months after platform entry will be analysed by the health economics team and is therefore not specified here.

2.9.4 Cost effectiveness

Similar to QoL at 12-months after platform entry we assume that this outcome will be analysed by the health economics team and is therefore not specified here.

2.9.5 All-cause mortality to 12 months

We will use a logistic regression model to analyse all-cause mortality to 12-months after platform entry. Specifically, if Y is a binary indicatoring with $Y = 1$ indicating death from any cause at 12-months, we model:

$$\begin{aligned} Y &\sim \text{Bernoulli}(p) \\ \text{logit}(p) = \eta &= x^\top \beta \end{aligned}$$

where η takes the same form as the primary analysis, see Section 2.6.

2.9.6 Clinical cure to 12 months

The occurrence of clinical cure, defined in the master protocol as no clinical or microbiological evidence of infection, is subject to the competing risk of death. We will use a logistic regression model to analyse clinical cure to 12-months after platform entry following an analogous approach to Section 2.9.5.

Given the possibility of death, this analysis actually yields a comparison of clinical cure (as defined within the context of the primary outcome) versus *either* clinical failure or death. That is, if the patient dies then we will consider them a failure, irrespective of whether they were clinically cured or not at the time of death.

2.9.7 No longer taking any antibiotics for the index joint to 12 months

The ability to observe whether antibiotics are being used for the index joint is subject to competing risks.

We will use a logistic regression model to analyse antibiotic use to 12-months after platform entry following an analogous approach to Section 2.9.5.

Given the possibility of competing risks, this analysis actually yields a comparison of no longer taking antibiotics (as defined within the context of the primary outcome) versus *either* taking antibiotics, joint removal, amputation or death. That is, if the patient dies then we will consider them a failure, irrespective of whether they were taking antibiotics or not at the time of death.

2.9.8 Destination prosthesis still in place to 12 months

The ability to observe whether the destination prosthesis is still in place is subject to the competing risk of death - we do not observe, counter to fact, whether those that die would or would not have had the destination prosthesis still in place.

We will use a logistic regression model to analyse destination prosthesis still in place to 12-months after platform entry following an analogous approach to Section 2.9.5.

Given the possibility of competing risks, this analysis actually yields a comparison of destination prosthesis still in place (as defined within the context of the primary outcome) versus *either* destination prosthesis no longer still in place or death. That is, if the patient dies then we will consider them a failure, irrespective of whether the destination prosthesis was or was not in place at the time of death.

2.9.9 Microbiological relapse b/w 100 days and 12 months

The occurrence of microbiological relapse, indicated by positive joint fluid or tissue culture for one or more of the index isolates, between 100 days and 12 months after platform entry, is subject to the competing risk of death. We will use a logistic regression model to analyse microbiological relapse to 12-months after platform entry following an analogous approach to Section 2.9.5.

Given the possibility of competing risks, this analysis actually yields a comparison of the presence of no microbiological relapse (with the prosthesis in place) versus *either* microbiological relapse or death (with or without the prosthesis in place). That is, if the patient dies then we will consider them a failure, irrespective of whether there had been a microbiological relapse or not at the time of death.

2.9.10 Microbiological reinfection b/w 100 days and 12 months

The occurrence of microbiological reinfection, indicated by positive joint fluid or tissue culture with a different organism to the index isolates, between 100 days and 12 months after platform entry, is subject to the competing risk of death. We will use a logistic regression model to analyse microbiological reinfection to 12-months after platform entry following an analogous approach to Section 2.9.5.

Given the possibility of competing risks, this analysis actually yields a comparison of the presence of no microbiological reinfection (with the prosthesis in place) versus *either* microbiological reinfection or death (with or without the prosthesis in place). That is, if the patient dies then we will consider them a failure, irrespective of whether there had been a microbiological reinfection or not at the time of death.

2.9.11 Time alive and free from revision procedure to 24 months

We will use a proportional-hazards model to analyse time alive and free from revision[1] to 24-months after platform entry. Specifically, assuming no time-varying covariates or coefficients:

[1]For the purposes of this outcome, we include amputation in the definition of revision.

$$h(t) = h_0(t)\exp\left(x^\top \beta\right)$$

where $h_0(t)$ is the baseline hazard at time t (weeks) and $x^\top \beta$ denotes the linear predictor. The baseline hazard will be modelled as a smooth function of time using splines, i.e. the above is modified such that

$$h(t) = \sum_{l=1}^{L} \gamma_l M_l(t; \vec{k}, \delta)\exp\left(x^\top \beta\right)$$

where $M_l(t; \vec{k}, \delta)$ denotes the l^{th} basis term for a degree δ M-spline function evaluated at knots locations $\vec{k}$ and γ_l denoting the l^{th} M-spline coefficient.

In the event of clear violation of proportional-hazards, appropriate model adjustments will be made as a sensitivity analysis. Alternatively, other model formulations may be explored, such as an accelerated failure time models under a log-logistic or Weibull distributional assumptions. Such models characterise effects in terms of survival times rather than hazards.

We will characterise treatment effects via the Restricted Mean Survival Time (RMST), defined as the area under the survival curve up to a specific time point. The RMST can be interpreted as the average time to event during the defined period to 24-months.

2.10 Supportive (domain specific) analyses

The following sections detail analyses that are domain-specific, i.e. where the interest is in treatment effects that are restricted to a given domain. For example, we are not interested in the effect of duration of antibiotic in the context of unplanned re-operation.

All these map to estimands are briefly outlined in the relevant DSAs.

2.10.1 Surgical domain

For the following, irrespective of the modelling approach, the effects of interest are only those relating to the surgical domain.

2.10.1.1 Treatment success at 12 months

todo - see notes

2.10.1.2 Unplanned re-operation

We will use logistic regression to analyse the occurrence of unplanned re-operation on the index joint more than 14 days after the initial definitive procedure. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.1.3 Dislocation of index joint

We will use logistic regression to analyse the occurrence of dislocation of the index joint on all patients irrespective of any occurrence of specified or unspecifed ICEs. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.1.4 Unplanned or unexpected periprosthetic fracture

We will use logistic regression to analyse the occurrence of unplanned or unexpected periprosthetic fracture (either intraoperative or later on, requiring attendance at a hospital). The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.2 Antibiotic duration domain

For the following, irrespective of the modelling approach, the effects of interest are only those relating to the antibiotic duration domain.

2.10.2.1 Treatment success at 12 months

todo - finalise approach for this alternative to itt.

2.10.2.2 Acute liver injury following platform entry to 90 days

We will use logistic regression to analyse the occurrence of acute liver injury to 90 days post platform entry. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.2.3 Laboratory-proven *Clostridium difficile* diarrhoea to x days

We will use logistic regression to analyse the occurrence of laboratory-proven *Clostridium difficile* diarrhoea. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.2.4 Antibiotics ceased due to suspected adverse reaction to 90 days

We will use logistic regression to analyse the occurrence of ceasing antibiotics due to other suspected adverse reaction between the time of domain entry to day 90. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.3 Extended prophylaxis domain

2.10.3.1 Treatment success at 12 months

todo - finalise approach for this alternative to itt.

2.10.3.2 Time alive and free from any revision procedure to 12-months

Time alive and free from any revision procedure on the index joint captured by a national joint replacement registry within 12 months of domain entry will be analysed via a proportional hazards survival model, using an analogous approach to that detailed in Section 2.9.11.

2.10.3.3 Time alive and free from any revision procedure to 24-months

Time alive and free from any revision procedure on the index joint captured by a national joint replacement registry within 24 months of domain entry will be analysed via a cause-specific hazard survival model, using an analogous approach to that detailed in Section 2.9.11.

2.10.3.4 Laboratory-proven *Clostridium difficile* diarrhoea to x days

We will use logistic regression to analyse the occurrence of laboratory-proven *Clostridium difficile* diarrhoea. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.3.5 Antibiotics ceased due to suspected adverse reaction to 90 days

We will use logistic regression to analyse ceasing antibiotics due to *other* suspected adverse reaction between the time of domain entry to day 90 post platform entry. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.4 Antibiotic choice domain

2.10.4.1 Treatment success at 12 months

todo - finalise approach for this alternative to itt.

2.10.4.2 Acute liver injury to day 100

We will use logistic regression to analyse the occurrence of acute liver injury to day 100 post platform entry. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.4.3 Acute liver injurty to day 100

2.10.4.4 Laboratory-proven *Clostridium difficile* diarrhoea to x days

We will use logistic regression to analyse the occurrence of laboratory-proven *Clostridium difficile* diarrhoea. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.10.4.5 Antibiotics ceased due to suspected adverse reaction to 100 days

We will use logistic regression to analyse ceasing antibiotics due to other suspected adverse reaction between the time of domain entry to day 100 post platform entry. The analyses will be based on all patients irrespective of any occurrence of specified or unspecifed ICEs. The outcome is subject to the competing risk of death.

2.11 Missing data

2.11.1 Primary outcome variable

Given the use of a composite outcome variable for the primary analysis, missingness for the different composite elements could lead to different implications on the interpretation of the outcome. The composite is ascertained through the use of hospital databases, followup with healthcare provider and/or patient and/or data linkage with death registries.

Treatment failure is indicated if any of the four composite elements are observed to fail. That is, if we see any of death, absence of clinical cure, ongoing antibiotic use or destination prosthesis absent, then the unit is a failure and whether other elements of the composite are observed or missing is irrelevant.

If direct evidence is missing of the patient being alive, this can be implied by the existence of any of observation of the other elements in the composite[2]. If all other elements are missing and death cannot be confirmed by other sources then the unit could either be alive or dead (e.g. they might have died out of country although this seems highly improbable). More likely

[2]Although being alive is only up until the time of the surrogate observation, which may or may not be prior to the 12 month endpoint.

is that if a patient cannot be found in a death registry, then they are alive (unless they moved overseas).

Clinical cure suggests that antibiotics would no longer be required and therefore could possibly be a proxy for antibiotic status if it were missing, although this would not be definitive. Conversely, if the status of clinical cure were missing and antibiotic status were known, then we could possibly take the fact that antibiotics were not being received to imply clinical cure.

Destination prosthesis status could be not be implied by other elements of the composite and therefore the outcome status cannot be confirmed if it is missing.

2.11.2 Covariates

When missing, some covariates can be completed by implication. For example, if region is missing, but site is available, region is known with certainty.

2.11.3 Handling missingness

When missingness is completely at random or dependent on covariates, but not jointly dependent on the exposure and the outcome nor dependent on an unobserved variable unrelated to the exposure, then, for logistic regression, the parameter estimate for the exposure is unbiased under complete case analysis (Hughes et al., 2019; McElreath, 2020, pg 503). As such, a complete case approach will be used at the interim analyses and final analysis as the headline inference. However, a sensitivity analysis will be run for the final analysis using multiple imputation that imputates both the missing components of the composite primary outcome variable and missing covariate values. Missing values will be imputed using a fully conditional specification via multivariate imputation by chained equations (Buuren and Groothuis-Oudshoorn, 2011).

2.12 Software

Analyses will be implemented in R and Bayesian models will be implemented in Stan and/or JAGS as required with posterior distributions approximated using Markov chain Monte Carlo. Also software used will be disclosed in the reports.

3 Quantities of interest

3.1 Treatment effects

Interest is primarily geared towards estimating effects of treatments under the specified analyses. In general, we consider randomised treatment whereby groups are exchangeable (i.e. the potential outcome associated with any treatment is independent of the treatment that was assigned). For the surgical domain, the choice of revision is left to the clinician and those groups are not exchangeable. This is because what leads a surgeon to prefer one versus two-stage revision will likely be informative of outcome, making those that receive one-stage systematically different from those that receive two.

The following descriptions relate to the calculation of treatment effects for the primary analysis. All domains adopt g-computation and a Bayesian bootstrap that aims to account for the uncertainty in the joint covariate distribution.

3.1.1 Surgical domain

For the surgical domain, we calculate the effect of revision relative to debridement using g-computation and a Bayesian bootstrap as follows:

- for each unit within the late acute silo, compute the predicted probability of the outcome assuming they are assigned to the DAIR group with each of their covariates set at their natural values, which, if necessary, are implied via reference to original allocation
- for each unit within the late acute silo and preferenced to one-stage revision, compute the predicted probability of the outcome assuming they are assigned to the revision group (and type determined from clinical preference) with each of their covariates set at their natural values, which, if necessary, are implied via reference to original allocation
- for each unit within the late acute silo and preferenced to two-stage revision, compute the predicted probability of the outcome assuming they are assigned to the revision group

(and type determined from clinical preference) with each of their covariates set at their natural values, which, if necessary, are implied via reference to original allocation
- for each set of predicted probabilities, form the scalar product with draws from a Dirichlet distribution of the relevant dimension to produce the average risk under DAIR, revision (one-stage), revision (two-stage) accounting for the uncertainty in the covariate distribution
- form a weighted combination of the the revision (one-stage), revision (two-stage) estimates based on the distribution of preferences in the sample
- derive the risk difference as the difference in the probability of the outcome in the revision versus DAIR groups

the difference between the sample level risk of treatment success under one-stage and two-stage revision minus the log-odds of treatment success under debridement.

3.1.2 Antibiotic duration domain

The antibiotic duration domain treatment effects are estimated using methods analogous to those in Section 3.1.1, restricting to the subset revealed to randomised antibiotic duration, all of which would have received a one-stage revision and who would therefore have not entered into the extended prophylaxis domain.

3.1.3 Extended prophylaxis domain

The extended prophylaxis domain treatment effects are estimated using methods analogous to those in Section 3.1.1, restricted to the subset revealed to randomised extended prophylaxis, all of which would have received a two-stage revision and who would therefore have not entered into the antibiotic duration domain.

3.1.4 Antibiotic choice domain

The antibiotic choice domain treatment effects are estimated using methods analogous to those in Section 3.1.1, restricted to the subset revealed to randomised antibiotic choice, with no restrictions on the entry into the other domains.

4 Decision procedures

4.1 Overview

Trial decisions are made by calculating the probability that the posterior distributions for the treatment effects exceed or fall short of reference values and contrast this with a decision level, all nominated based on simulation studies and the clinical subject matter knowledge.

The trial considers:

- superiority - indicating that an intervention is beneficial relative to a comparator
- non-inferiority - indicating that an intervention is no worse than some clinically relevant threshold to a comparator
- futility - which indicates a low probability with respect to superiority or non-inferiority assessment and thus there are both futility for superiority and futility for non-inferiority assessments.

The mathematical definitions of these quantities are provided below and a table of the thresholds currently in use is defined in Table 4.1.

Table 4.1: Decision thresholds

Domain	Quantity	Threshold probability	Decision probability
Surgical	Superiority revision vs DAIR	> 0	≥ 0.92
	Futility revision vs DAIR	> 0.05	≤ 0.3
Antibiotic duration	Non-inferiority 6 weeks vs 12 weeks	> -0.05	≥ 0.925
	Futility 6 weeks vs 12 weeks	< 0	≤ 0.25
Extended prophylaxis	Superiority 12 weeks vs none	> 0	≥ 0.95
	Futility 12 weeks vs none	> 0.05	≤ 0.25
Antibiotic choice	Superiority rifampacin vs none	> 0	≥ 0.995
	Futility rifampacin vs none	> 0.05	≤ 0.25

The actions that follow after a decision threshold has been reached are detailed in the core protocol (section 9.7) and will not be duplicated here. However, any occurrence of a futility decision requires that a subgroup analysis be performed as per the populations listed in Section 2.2 using the methods detailed in Section 2.8 for the relevant domain(s).

For a futility decision to hold definitively, all relevant subgroups should show futility. Variations to this are permitted at the DSMC disgression.

4.2 Superiority

Superiority implies that an intervention has high probability of being beneficial relative to a comparator intervention. Superiority is a posterior probability assessment against a threshold value (for the current domain threshold values see Table 4.1).

$$\Pr(\Delta > \rho_{sup}) \geq \gamma_{sup}$$

as a boolean indicator of treatment superiority based on the posterior estimate treatment comparison represented by Δ relative to a reference value ρ_{sup} at a probability greater than or equal

to a threshold value γ_{sup}.

4.3 Non-inferiority

Non-inferiority implies that an intervention has a high probability of being no worse than some pre-specified threshold relative to the comparator intervention.

$$\Pr(\Delta > \rho_{ni}) \geq \gamma_{ni}$$

as a boolean indicator of treatment non-inferiority based on the posterior estimate treatment comparison represented by Δ relative to a reference value ρ_{ni} at a probability greater than or equal to a threshold value γ_{ni}.

4.4 Futility

Futility in relation to either a superiority or a non-inferiority assessment implies that an intervention has low probability relative to the comparator intervention with respect to that assessment.

$$\Pr(\Delta > \rho_{fut}) \leq \gamma_{fut}$$

as a boolean indicator of treatment futility based on the posterior estimate treatment comparison represented by Δ relative to a reference value ρ_{fut} at a probability greater than or equal to a threshold value γ_{fut}. Separate futility assessments are run for both superiority and non-inferiority.

5 Adaptation

5.1 Adaption considerations

Three possible adaptations are currently identified, namely early stopping (due to superiority, non-inferiority or futility), the addition of interventions within existing domains and the addition of entirely new domains.

6 References

Buuren, S. van, Groothuis-Oudshoorn, K., 2011. Mice: Multivariate imputation by chained equations in r. Journal of Statistical Software 45, 1–67. https://doi.org/10.18637/jss.v045.i03

De Schryver, M., 2019. A tutorial on probabilistic index models: Regression models for the effect size p(Y1 > Y2). American Psychological Association 24. https://doi.org/10.1037/met0000194

Evans, S., Follmann, D., 2016. Using outcomes to analyze patients rather than patients to analyze outcomes. Stat Biopharm Res 8. https://doi.org/10.1080/19466315.2016.1207561

Hughes, R., Heron, J., Sterne, J., Tilling, K., 2019. Accounting for missing data in statistical analyses: Multiple imputation is not always the answer. International Journal of Epidemiology 48, 1294–1304. https://doi.org/10.1093/ije/dyz032

Mallinckrodt, C., Molenberghs, G., Lipkovich, I., Ratitch, B., 2020. Estimands, estimators and sensitivity analysis in clinical trials. CRC Press, London.

McElreath, R., 2020. Statistical rethinking. CRC Press.

Rosenberger, W., Lachin, J., 2016. Randomization in clinical trials. Wiley.