Statistical Analysis Plan

ROADMAP: RandOmised Arthroplasty infection worlDwide Multidomain Adaptive Platform trial

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

Version: 0.1

Date: 11 October 2024

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	9/2024	First version	N/A

Repository status

```
## Local:   main /Users/mark/Documents/project/roadmap/src/roadmap-sap
## Head:     [8858df4] 2024-10-03: wip
##
## Branches:         1
## Tags:             0
## Commits:         11
## Contributors:     1
## Stashes:          0
## Ignored files:    3
## Untracked files:  7
## Unstaged files:   2
## Staged files:     0
##
## Latest commits:
## [8858df4] 2024-10-03: wip
## [5054c33] 2024-10-01: Update to fix repeat row name in table for gt tables
## [8789fea] 2024-10-01: Move to gt for tables
## [cb17729] 2024-09-19: wip
## [2135b2e] 2024-09-18: wip
```

1 Introduction

ROADMAP is a platform study for pragmatically evaluating the efficacy of multi-modal interventions used in the treatment of prosthetic joint infection (PJI). The primarily 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 overlapping reports and thereby 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[1], specifically:

1. For *late-acute* 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 after *one-stage revision* surgery?
3. Is it better to include rifampicin in the backbone antibiotic treatment regimen?

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

1. Alive
2. Clinical cure (no microbiological evidence of infection)

[1]Italics in the bullet list indicate updates that might be required for protocol.

3. No ongoing use of antibiotics for the index joint
4. Destination prosthesis in place after the initial management strategy is complete

The absence of any of the above elements is considered to be a treatment failure for the purposes here. More detail on the primary outcome definition is given in the protocol.

1.2 Objectives and endpoints

The overarching objective for ROADMAP trial is to examine the effect of a range of interventions in patients with PJI (defined by the eligibility criteria) on treatment success where that is defined as the patient being alive with clinical cure, not taking antibiotics for PJI and with the destination prothesis in place and is assessed at 12 months[2]. Table 1.1 provides a high-level summary of the domain objectives. Further detail can be found in the protocol.

Table 1.1: ROADMAP domain-level analytic objectives

Domain	Objective	Endpoint
Surgical	Determine whether revision (one or two-stage) is more effective than debridement in curing prosthetic joint infections in late-silo units.	Treatment success at 12 months (alive, clinical cure, 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 ext-prophylaxis is more effective than none in curing 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 in curing 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.

[2]Repitition intentional.

1.3 Estimands

The estimands are detailed in the core protocol (references below correspond to those in the protocol) and also in the domain specific appendicies.

> **Note**
>
> The estimands are introduced in generality in the core protocol and then made specific in the DSA but not much more detail was added. We might be better served by removing estimands from the core protocol and retaining their definitions in the DSAs. They need more detail than they presently have but we can also flesh out further here. Mallinckrodt et al. (2020) gives guidanace on how to improve the current specification. The way the estimands are presently stated is tokenistic, they don't provide any real value.

Estimands emphasise what is being estimated rather than how the estimation is undertaken. Below we add detail to the specification provided in the protocol and DSAs with a focus on the treatment regimen, population summaries and intercurrent events.

> **Note**
>
> Notice that each of the treatment-specific ICEs included in the discussion that follows may occur in isolation for a given unit or in conjunction with ICEs associated with treatment arms in other domains. For example, a patient switches from DAIR to revision due to having a loose joint and then also discontinues extended prophylaxis due to adverse reaction. In such instances, the ICEs are still handled by the applicable strategy for the estimand under consideration. This implies, for the above example, that for the surgical domain comparison we would likely consider the implication of the cross-over alone and not contemplate the impact that discontinuing may have. Similarly, when considering the extended prophylaxis comparison, we would likely only refer to the ICE associated with discontinuation and not the ICE relating to the loose joint.

1.3.1 Primary estimand

For each domain there is a primary estimand for which a generic overview is given in the core protocol and then a specific version provided in the DSAs.

The primary estimand for all the domains adopts an intention to treat (ITT) perspective in that all ICEs are to be handled under a treatment policy strategy. This means that all ICEs are effectively considered to be part of the treatment regimen of interest and are therefore be ignored, Mallinckrodt et al. (2020). However, the definition of the primary outcome does, in fact, imply a composite strategy for (what would be) ICEs relating to both death and the replacement of desination prothesis.

The detail provided here is intended to clarify intent and aid thinking and interpretation as the use of a ICE strategies have specific implications that would otherwise remain unstated. All the identifiers relate to those used in the DSAs.

> **Note**
>
> 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. This has to be kept in mind when interpreting and generalising the results.

1.3.1.1 Estimand B.1 (surgical intervention)

High-level lay overview:

The estimand considers the effect of removing the infected joint versus cleaning and leaving the joint in place on treatment success at 12 months.

ICEs/Treatment regimen:

The protocol definitions for the surgical interventions are given in the DSA for the Surgical strategy domain (late acute silo). These definitions provide detail for all the 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 etc.

The randomised treatment arms comprise DAIR and revision with the clinician permitted to self-select the type of revision (one or two-stage) to be used when a patient is assigned to the revision group. The treatment regimen being evaluated is expanded upon below.

Likely intercurrent events for the surgical domain and the approaches used to handle them are summarised in Table 1.2.

> **Note**
>
> Other ICEs are possibly relevant but not yet considered. For example, withdrawal due to the clinical opinion on what is in the patient's best interest with regards to their involvement in roadmap. However, even if a patient is withdrawn by request of the clinician, it may be acceptable to continue to collect data on such a patient, in which case the treatment policy would still be relevant.
> Deviations from assigned treatment in terms of variations in procedures per the protocol definitions but not leading to switching would also constitute an ICE that impact the interpretation of the treatment effects. An example of this may be secondary surgery that does not lead to replacement of prosthetic, i.e. not handled via the composite strategy. Another example may be deviation from any protocol timings such as a requirement to have a surgical procedure within a given time window from randomisation. Within this estimand, these would also be currently handled via the treatment policy strategy.

Table 1.2: Surgery domain intercurrent events

ID	Objective	Endpoint	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.	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.	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.	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.

The treatment regimen under evaluation is the randomised and revealed surgical 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. As such, the estimand is not associated with a specific single treatment protocol, but rather the average of a total effect of a mix of regimens aligned to those that commonly occur in practice when undertaking DAIR or revision.

> **Note**
>
> In the above, perhaps it would be more accurate to say that the regimen is the "randomised and revealed surgical treatment ... in additon to a specific combination of randomised and non-randomised treatments associated with other domains."

Population summary:

The population-level summary measure is the conditional odds ratio for revision comprising both one-stage and two-stage procedures relative to DAIR.

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.

Main estimator:

A single Bayesian multivariable logistic regression model is used to characterise unit responses for the primary outcome across all domains. Further detail provided in later sections.

Given the structure of the study, specifically, the desire to permit clinicians free choice of revision type and the dependency between specific revision types and entry into the antibiotic duration and extended prophylaxis domain, the population-level summary is constructed as a weighted conditional log odds ratio where the weights adopted are to be informed by the sample distribution for the surgical revision type under assignment to revision.

Missing data:

Missing data is inevitable due to, for example, limitations in identifying clinical records, formal requests to completely withdraw from the study (including use of data), loss to follow up due to relocation 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). However, missingness is assumed to be low and also at random across all variables and variable components such that imputation could, in principle, be undertaking by conditioning on observed variates.

Under the above assumptions, that is 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. In this setting, a complete-case (likelihood-based) analysis will yield unbiased estimates of the treatment effects. Therefore, 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 being disclosed in the reporting.

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 lay overview:

The estimand considers the effect of the duration of backbone antibiotics on treatment success at 12 months, specifically whether 6 weeks of backbone antibiotic is non-inferior to 12 weeks in patients receiving one-stage revision.

ICEs/Treatment regimen:

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

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	Objective	Endpoint	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.	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.	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 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.

> **Note**
>
> The more I review this, the more I wonder if it really is an accurate representation of the regimen of interest and whether our single model (highly conditional) perspective is actually what is required. This is despite the fact that I believe that the conditional perspective is the only one that really makes sense here in terms of being reasonably well defined and that the inherently vague pragmatic perspective is not the right way to approach this research problem.

Population summary:

The population-level summary measure is the conditional odds ratio for 6 weeks backbone antibiotic relative to 12 weeks.

Other sections are as defined for B.1.

1.3.1.3 Estimand E.1 (extended prophylaxis intervention)

High-level lay 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 of the revision procedure in patients receiving two-stage revision.

ICEs/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.

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	Objective	Endpoint	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.	Extension to therapy[1]	Treatment policy
E.1(i2)	None	If the patient shows late positive culture, they may become ineligible after revealed to their and would also receive an additional 6-12 weeks of backbone antibiotic.	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.	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.

> **Note**
>
> todo - Get more detail on E.1(i2) and check my interpretation is correct.

The treatment regimen under evaluation is 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.

Population summary:

The population-level summary measure is the conditional odds ratio for 12 weeks extended prophylaxis relative to none.

Other sections are as defined for B.1.

1.3.1.4 Estimand C.1 (antibiotic choice intervention)

High-level lay 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 none for domain eligible units.

ICEs/Treatment regimen:

The protocol definitions of the interventions are provided in the DSA for the Antibiotic choice domain.

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	Objective	Endpoint	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.	Switchover	Treatment policy
C.1(i2)	Rifampicin	If the site pharmacy runs out of or does not have any rifampicin in stock may lead to early termination or non-receipt of therapy.	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.	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.

> **Note**
>
> Clarify C.1(i1) - would this still constitute switchover after the delayed start (or rescue)? As with other domains, patient discharge/relocation is probably relevant as another ICE.

The treatment regimen under evaluation is 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.

Population summary:

The population-level summary measure is the conditional odds ratio for rifampicin relative to none.

Other sections are as defined 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.

By and large, all of the supportive estimands should be treated as exploratory.

1.3.2.1 Estimand B.2 (surgical intervention)

High-level lay overview:

Per the primary estimand for the surgical domain, this secondary/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.

> **Note**
>
> The manner in which the estimand is currently written, aligns with a per-protocol analysis whereby only those patients that were adherent to their assigned (observed) treatment are retained in the analysis. This doesn't consider whether those patients would also be adherent in their counterfactual treatment. The problem with this is that it violates the randomisation - the groups might be unbalanced on either known or unknown (or both) confounders. The traditional per-protocol effect is well known to produce biased estimates for treatment effects. What I believe is of interest is the effect of the surgical intervention in those that would have adhered to all arms. Given that the ICEs for the surgical domain negate the possibility of forcing the assignment on the patient, principal strata seems like the only approach that can be used. I am aware that views differ on this

and we need to discuss.

ICEs/Treatment regimen:

The only ICEs are per the primary estimand for the surgical domain.

> **Note**
>
> The DSA currently states that principle stratum strategy is to be used as the single mechanism for handling ICEs. Obviously, this is a placeholder as it was a fairly arbitrary inclusion. For the surgical domain currently specified ICEs are switchover to revision for DAIR where the joint is loose and similarly for revision under opposite conditions. Additionally, revision might be abandoned if the patient becomes unstable during the operation. To handle these via desired strategy we first conceptualise strata. The subset associated with the compliers (however that is defined here) would be the one for which we would estimate the effect of DAIR vs revision.

Population summary:

todo

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)

todo

1.3.2.3 Estimand E.2 (extended prohylaxis)

todo

1.3.2.4 Estimand C.2 (antibiotic choice)

todo

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 efficacy 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 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 subject to steering committee and ethics review.

ROADMAP will be conducted sequentially (cohorts of 500 patients) with decision rules evaluated on parameter estimates of interest, driving 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 inherent flexibility, ease of both uncertainty quantification and their capacity for incorparating adaptive elements, regularisation of parameter estimates and simplicity in interpretation.

1.4.1 Randomisation

Units will be randomised (under fixed complete randomisation, i.e. non-adaptive and without restriction through blocking or stratification etc) to one arm within each domain for which they are eligible (Rosenberger and Lachin, 2016). This method was selected for its operational simplicity. As entry into the duration domains are dependent on the status of the surgical

intervention, some subtleties arise. For example, only units in receipt of one-stage revision can be revealed to randomised assignment for antibiotic duration which also prevents entry into extended prophylaxis. Similarly, only units in receipt of two-stage revision can be revealed to randomised assignment for extended prophylaxis. A practical way to operationalise this is to simply randomise each unit to all available domains and then reveal the allocation as necessary and that is the approach used here.

As a concrete example, a unit within any of the early, late or chronic silo would be allocated interventions for all domains. However, for early silo units, the surgical interventions are irrelevant and so never revealed, the antibiotic duration interventions would only be revealed if one-stage was received, the extended prophylaxis would only be revealed if two-stage was received and so on.

1.4.2 Sample size

The initial trial funding and infrastructure has sufficient resources to enrol up to 2,500 participants into the platform. Thus, the sample size is initially constrained by the current resources and the desired structure. However, the study is intended to be perpetual and will seek funding to ensure recruitment beyond the initial sample size if necessary. Implications of the initial sample size on the amount of uncertainty that can be resolved under assumptions about the cohort population, the present design and assumed effects have been evaluated by simulation, discussed later.

1.4.3 Data management

An overview is provided in the Master Protocol. However, we note that the data storage approach will be decomposed into redcap components and (out-sourced proprietary) platform components developed by Spiral Software[3], the latter also being responsible for the implementation of the randomisation processes.

> **Note**
>
> todo - double check and write up on how this functionality is being decomposed. What are the redcap parts and what belongs to spiral?

[3] https://spiral.co.nz/

2 Statistical methods

2.1 Analysis sets

2.1.1 Intention to treat

The intent to treat (ITT) principle covers what units to include and what data to include on each unit. A strict interpretation of ITT demands collection and analysis of all data, but minor deviations from this are routinely accepted.

Estimands B.1 through C1 are intended to align with the ITT perspective via the treatment policy strategy. The analysis population used for these will include all units that were randomised and revealed to at least one of the domain-level interventions and have passed the primary endpoint of 12-months with their primary endpoint status known or known to be missing.

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 until the data has been entered. Participants who have been randomised, but have not yet reached the primary endpoint, will be excluded.

2.1.2 Per-protocol

The minimal requirements for determining that patients are aligned with the protocol criteria with regards to the manner in which the interventions were delivered are provided in Table 2.1.

> **Note**
>
> Per-protocol analyses traditionally exclude those patients that deviate from the protocol and conduct the primary analysis with this subset. However, the results from such analyses will only be valid where deviations from the protocol are non-informative. Consider (1) exclusions can imbalance treatment groups on known or unknown confounders and prognostic factors (2) that noncompliance may relate to a specific intervention or disease severity producing differential dropout across arms.

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	5 to 7 weeks of antibiotic therapy has been completed, between the date of the one-stage revision and 16 weeks later
	12 weeks	11 to 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 effects to subgroup populations enable effect heterogeneity to be explored, but come at the increased risk of instability, bias and false positives. Subgroup populations have been identified in the various appendices but are duplicated here in summary form for convenience, Table 2.2.

Table 2.2: Subgroup populations

Domain	Subgroup
Surgical	Site of infection by joint (hip/knee)
	Duration of symptoms at domain entry in 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

2.3 Description

A CONSORT diagram will will be provided to detail patient progression showing:

- participants screened
- participants eligible (giving reasons for ineligibility)
- participants consented

- participants entering into randomised treatment (revealed) by domain and intervention
- participants withdrawing from study
- participants reaching 12-month endpoint

Recruitment numbers will be reported by region and site and intervention availability by domain will be presented, again by site.

Protocol deviations and the occurrence of intercurrent events will be summarised by domain and intervention. Baseline demographics will be provided by domain and intervention.

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 the primary endpoint and every 4 months thereafter.

> **Note**
>
> todo - revisit this timing. I think it is probably more operationally practical to run interims on a regular timing rather than for fixed sample size.

The analysis will focus solely on the primary outcome (with subgroup analyses run conditional on futility rules being triggered, see later) and will follow the approach detailed in Section 2.6. Notwithstanding the clinicial team and patients involved in delivery and receipt of the interventions, only the analytical and data groups responsible for providing analysis results to the DSMC will be privy to individual treatment group assignments.

2.5 Analysis approach

Analyses will be conducted within a Bayesian framework with a focus on the estimation of estimands as detailed earlier. 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. 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, region, site, time period of recruitment and baseline characteristics, adopting the following form:

$$
\begin{aligned}
Y &\sim \text{Bernoulli}(p) \\
\text{logit}(p) &= \mu + \lambda_s + \rho_j + \phi_q + \sum_{d \in \mathscr{D}} \vec{\mathbf{x}}_d' \vec{\beta}_d + \tau_t + \psi_r + \zeta_{z(r)} + \vec{\mathbf{w}}' \vec{\gamma}
\end{aligned}
$$

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

- μ 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 preference for type of revision surgery q
- $\vec{\beta}_d$ shift attributable to each vector of treatment effects (including non-participation effects) as indicated by the row vector $\vec{\mathbf{x}}'$ for domain d in the set of available domains $\mathscr{D}$
- ψ_r shift attributable to region r
- $\zeta_{z(r)}$ shift attributable to site z nested within region r
- τ_t shift attributable to randomisation period t
- $\vec{\gamma}$ shift attributable to baseline characteristics as indicated by the row vector $\vec{\mathbf{w}}$

Constraints, as necessary, will be imposed for identifiability and disclosed in the reporting.

In order to estimate the treatment effects of interest we use g-computation as described in Section 3. Parameters estimates will be reported as point and interval summaries of the posterior.

2.6.1 Priors

In general we will use weakly informative priors that will not overly constrain the results, but will be consistent with the belief that extreme treatment effects are unlikely.

The prior for μ will be set to

$$\mu \sim \text{Logistic}(0, 1)$$

which is centred on zero with 90% of its mass between ±3 on the log-odds scale and uniform on the probability scale.

The priors for all main effects will be set to

$$(\lambda_s, \rho_j, \phi_q, \vec{\beta}_d, \vec{\gamma}) \sim \text{Normal}(0, 2.5)$$

which are centred on zero with 90% of its mass between ±3.3 on the log-odds scale.

The baseline covariates in the model include

- age
- sex
- ...

Noting that some domains may vary their active set of interventions over time, a first-order random walk will be used to model temporal variation in the background response which 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 the τ_1 term representing the most recent analysis.

> **Note**
>
> Question - Would it be preferable/more sensible to model on a finer granularity, i.e. monthly, quarterly etc?

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

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

The primary analysis model will be assessed for adequacy. Additional models (either simpler or more complex) may be investigated as part of checks of sensitivity, stability, and model fit. If any issues or concerns arise (for example, strong evidence of interactions across treatment domains), all changes or updates to the specified primary model will be documented and reported.

2.7 Sensitivity analysis (applicable to primary)

2.8 Subgroup analyses

Pre-specified subgroup analyses will be restricted to the primary estimand with additional post-hoc exploratory subgroup analyses being discretionary. The general approach will be to use the complete data, incorporating first-order interactions within the relevant models via hierarchical models for each subgroup considered. Analyses will be run for final reporting on 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. Additionally, the researchers had 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. For example, for evaluation of silo-specific effects, the revised linear predictor would take the following form:

$$\mu_s + \rho_{j,s} + \phi_{q,s} + \sum_{d \in \mathscr{D}} \vec{\mathbf{x}}_d' \vec{\beta}_d + \tau_t + \psi_r + \zeta_{z(r)} + \vec{\mathbf{w}}_s' \vec{\gamma}_s$$

leaving time and site effects unstratified (along with terms where identification is infeasible). For each of the main effects, priors are converted to a hierarchical structure, for example for the surgical domain treatment effects:

$$
\begin{aligned}
(\vec{\beta}_{1,s}) &\sim \text{Normal}(\nu, \sigma_\nu) \\
\nu &\sim \text{Normal}(0, 2.5) \\
\sigma_\nu &\sim \text{Half-Normal}(0, 1)
\end{aligned}
$$

and with other main effects dealt with analogously.

2.9 Supportive (domain agnostic) analyses

The following sections detail analyses applying to all domains, i.e. where interest is in treatment effects associated with all domains.

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) 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.

> **Note**
>
> Care has to be taken when interpreting a PI. First, the measure is NOT the probability that a patient will benefit from having the experimental treatment rather than control. For further cautions, see Senn (2011).

Separate DOOR analyses will be run on strata based on the assessment and ranking of the relevant outcomes at 12-months:

- 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

> **Note**
>
> We note that a Bayesian approach to the MWU exists but has not been widely adopted, Chechile (2020).

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. As of 2007, 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 (no, 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 proportional-hazards model to analyse all-cause mortality to 12-months after platform entry. Specifically, assuming no time-varying covariates or coefficients:

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

where $h_0(t)$ is the baseline hazard at time t (weeks) and $x^\top \beta$ denotes the linear predictor. he baseline hazard 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.

> **Note**
>
> The M-spline should include an intercept so that $h(t)$ is not constrained to zero at the origin and the coefficients are constrained to a simplex for identifiability. todo - go back and review M-spline vs B-spline differences in approach, stability and so on.

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 therefore use a cause-specific proportional-hazards model to analyse this outcome. Specifically, assuming no time-varying covariates or coefficients, for risk r we have:

$$h_{r(t)} = h_{0,r}(t) \exp\left(x^\top \beta_r\right)$$

This method evaluates the intensity associated with a given event type, treating all competing events as censored, in addition to those who are censored for loss to follow-up or withdrawal.

The approach assumes independent censoring for patients who are not truly censored but that fail fail from the competing event, most commonly death. This means that any censored patient at time t would have the same rate of clinical cure, regardless of whether the censoring was due to administrative censoring or any of the other event types. There is no way to explicitly test this assumption.

> **Note**
>
> todo - would a multi-state model perspective be more appropriate for these?

2.9.7 No longer taking any antibiotics for the index joint

We will use a logistic regression model to analyse the proportion no longer taking antibiotics at 12-months. Specifically, if Y is a binary indicatoring with $Y = 1$ indicating use of antibiotics 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.

Given the possibility of death, this analysis actually yields a comparison of no longer taking antibiotics (as defined within the context of the primary outcome) versus *either* taking antibiotics or death.

2.9.8 Destination prosthesis still in place

We will use a logistic regression model to analyse the proportion with the destination prosthesis still in place at 12-months after platform entry.

Similar to the analysis defined in Section 2.9.7, this is actually a comparison of having the destination prosthesis in place (as defined in the primary outcome) versus *either* not having the destination prosthesis in place or 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.

Therefore, this will be analysed via a cause-specific hazard survival model, analogous approach as that detailed in Section 2.9.6.

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.

Therefore, this will be analysed via a cause-specific hazard survival model, analogous approach as that detailed in Section 2.9.6.

2.9.11 Time to joint registry captured revision procedure to 24 months

> **Note**
>
> todo - Clarify what the intent for this is.

2.10 Supportive (domain specific) analyses

The following sections detail analyses are domain-specific, i.e. where the interest is in treatment effects specific 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.

> **Note**
>
> todo - Get the DSA estimand specifications update to indicate that these are all up to 12 months.

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

> **Note**
>
> todo - finalise approach for this alternative to itt. This is within the domain-specific section as we take a domain specific approach for the ICEs. Ditto with the other domains. Discuss.

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.

> **Note**
>
> todo - Check whether this was meant to be 90 days **post commencement of antibiotics** rather than post platform entry, the former requiring a different approach.

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.

> **Note**
>
> todo - Check whether this should have a time limit on it. Presumably also 90 days...

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.

> **Note**
>
> todo - What is meant by other? Check whether this was meant to be 90 days post commencement of antibiotics.

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 is subject to the competing risk of death. Therefore, this will be analysed via a cause-specific hazard survival model, analogous approach as that detailed in Section 2.9.6.

> **Note**
>
> todo - This seems like it would be domain agnostic. Not sure what it is doing in the DSA for ExP

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 is subject to the competing risk of death. Therefore, this will be analysed via a cause-specific hazard survival model, analogous approach as that detailed in Section 2.9.6.

> **Note**
>
> todo - Again, this seems like it would be domain agnostic. Not sure what it is doing in the DSA for ExP

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.

> **Note**
>
> todo - Check whether this should have a time limit on it. Presumably 90 days as per ab duration domain

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.

> **Note**
>
> todo - What is meant by other? Check whether this was meant to be 90 days post commencement of extended prophylaxis.

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 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.

> **Note**
>
> todo - Check whether this should be **post commencement of rif**.

2.10.4.3 Acute liver injurty (to day 100)

> **Note**
>
> todo - Check if c.4 is a duplicate??

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.

> **Note**
>
> todo - Check whether this should have a time limit on it. Presumably 90 days as per ab duration domain

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.

> **Note**
>
> todo - Check why this is 100 days and others are 90.

2.11 Missing data

2.11.1 Primary outcome variable

Given the use of a composite outcome variable for the primary analysis, missingness in the composite elements can have differing implications to 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.

Being alive, if direct evidence is missing, can be implied by the existence of any of observation of the other elements in the composite. 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 is that if a patient cannot be found in a death registry, then they are alive.

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 then, for logistic regression, the parameter estimate for the exposure is unbiased under complete case analysis (Hughes et al., 2019). 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 the conditional effects of treatments under the various analyses. In general, treatment is randomised so 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.

> **Note**
>
> To expand on the above, assume that when a unit is randomised to revision and the clinical decision of what type of revision to do is based solely on disease severity with more severe units receiving one-stage and those that are less severe receive two-stage. This violates the exchangeability assumption in that if $\mathbb{E}[Y|T=1] = y_1$ and $\mathbb{E}[Y|T=2] = y_2$ (with Y denoting the observed outcomes and $T=1$, $T=2$ denoting receipt of one and two-stage) then we would not obtain the same expectations if the sets of treatment units were switched. In other words, $\mathbb{E}[Y(1)|T=1] \neq \mathbb{E}[Y(1)|T=2] \neq \mathbb{E}[Y(1)]$ and $\mathbb{E}[Y(2)|T=1] \neq \mathbb{E}[Y(2)|T=2] \neq \mathbb{E}[Y(2)]$. To remedy this, we adjust for a covariate set (in this example, disease severity but in the primary analysis we adjust for revision type preference and additional patient characteristics) with the goal of achieving conditional exchangeability.

The following descriptions relate to the calculation of treatment effects (conditional log-odds-ratios) for the primary analysis. All domains using g-computation and a Bayesian bootstrap to characterise the uncertainty in the covariate distribution.

3.1.1 Surgical domain

For the surgical domain, we calculate the effect of revision relative to debridement as the difference between a weighted sum of the log-odds of treatment success under one-stage and two-stage revision minus the log-odds of treatment success under debridement.

Specifically, we approximate the following integrals based on the fitted values for the patients in the late-acute silo:

$$\begin{aligned}
\mathbb{E}[\hat{\eta}(d_1 = 1, d_2 = 1, d_3 = 1, l)] &= \int_{\mathscr{L}} \hat{\eta}(d_1 = 1, d_2 = 1, d_3 = 1, l) d\hat{P}_L(l) \\
\mathbb{E}[\hat{\eta}(d_1 = 2, d_2 = 1, d_3 = 1, r = 1, l)] &= \int_{\mathscr{L}} \hat{\eta}(d_1 = 2, d_2 = 1, d_3 = 1, r = 1, l) d\hat{P}_L(l) \\
\mathbb{E}[\hat{\eta}(d_1 = 3, d_2 = 1, d_3 = 1, r = 2, l)] &= \int_{\mathscr{L}} \hat{\eta}(d_1 = 3, d_2 = 1, d_3 = 1, r = 2, l) d\hat{P}_L(l)
\end{aligned}$$

where $\hat{\eta}$ denotes the linear predictor function, d_k denotes the interventions in the surgical ($k = 1$), duration ($k = 2$) and extended prophylaxis domains ($k = 3$) with the surgical domain (d_1) indexes corresponding to debridement, one-stage and two-stage respectively. Preference for surgical type under revision is denoted by r and, finally, l denotes the remaining covariate set including preference where it is not explicitly conditioned on and **other domain level interventions** not fixed for the purposes of the comparison, site, time effects etc. We compute the effect of revision relative to debridement as:

$$\begin{aligned}
\Delta_{d_1[1]} = \mathbb{E}[\hat{\eta}(d_1 = 2, d_2 = 1, d_3 = 1, r = 1, l)]\mathbb{E}[r = 1] + \\
\mathbb{E}[\hat{\eta}(d_1 = 3, d_2 = 1, d_3 = 1, r = 2, l)]\mathbb{E}[r = 2] - \\
\mathbb{E}[\hat{\eta}(d_1 = 1, d_2 = 1, d_3 = 1, l)]
\end{aligned}$$

where $\Delta_{d_1[1]}$ denotes the log-odds-ratio comparison of interest[1], $\mathbb{E}[r = 1]$ and $\mathbb{E}[r = 2]$ are the relevant weights based on the observed distribution of preference for revision type in the sample.

[1]There is currently only one comparison of interest for each domain but the index gives us the ability to refer to others should the need arise.

> **Note**
>
> Question - Would it be more sensible to compute $\mathbb{E}[Y|\dots]$ from the inverse link transformed linear predictor and then compute the log-odds ratios from there or is the above acceptable???

3.1.2 Antibiotic duration domain

The antibiotic duration domain treatment effects are simpler to estimate than those in the surgical domain in that we can calculate the effect of 6 weeks of backbone antibiotic relative to 12 weeks directly. To do so, we use methods analogous to those detailed above 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

For the extended prophylaxis domain we calculate the effect of 12 weeks of extended prophylaxis relative to none again using methods analogous to those detailed above. The estimation procedure is 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

For the antibiotic choice domain we calculate the effect of the addition of rifampacin relative to none again using methods analogous to those detailed above. The estimation process is 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 based on probabilities associated with the comparisons of interest as characterised by the model parameters discussed earlier. In its current format, 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 - indicating a low probability with respect to a superiority or non-inferiority assessment, as applicable

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

	Quantity	Threshold	Probability
Surgical	Superiority revision vs debridement	> 0	≥ 0.99
	Futility revision vs debridement	$> \log(1.2)$	≤ 0.05
Antibiotic duration	Non-inferiority 6 weeks vs 12 weeks	$> \log(1/1.2)$	≥ 0.99
	Futility 6 weeks vs 12 weeks	$< \log(1/1.2)$	≤ 0.2
Extended prophylaxis	Superiority 12 weeks vs none	> 0	≥ 0.99
	Futility 12 weeks vs none	$> \log(1.2)$	≤ 0.05
Antibiotic choice	Superiority rifampacin vs none	> 0	≥ 0.99
	Futility rifampacin vs none	$> \log(1.2)$	≤ 0.05

The actions that occur when decision threshold are met have been detailed in the core protocol and will not be duplicated here. However, any occurrence of a futility decision requires 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, it is expected that all relevant subgroups would show futility.

4.2 Superiority

Superiority implies that an intervention has high probability of being beneficial relative to a comparator intervention. Generally, superiority is a probability assessment against a zero reference. For any of the treatment effect parameters obtained from the primary analysis results we define:

$$\mathbb{I}^{\text{sup}}_{d_k[c]} = \Pr(\Delta_{d_k[c]} > s_{d_k[c]}) \geq \gamma^{\text{sup}}_{d_k[c]}$$

as treatment superiority where $\mathbb{I}^{\text{sup}}_{d_k[c]}$ is a boolean indicator of superiority, for the comparison indexed by c within the d_k domain relative to a reference level $s_{d_k[c]}$ (generally zero) at a probability greater than or equal to $\gamma^{\text{sup}}_{d_k[c]}$.

4.3 Non-inferiority

Non-inferiority implies that an intervention has a high probability of being no worse than some pre-specified clinical threshold relative to the comparator intervention. Generally, non-inferiority is a probability assessment against $\log(1/1.2)$. For any of the treatment effect parameters obtained from the primary analysis results we define:

$$\mathbb{I}^{\text{ni}}_{d_k[c]} = \Pr(\Delta_{d_k[c]} > n_{d_k[c]}) \geq \gamma^{\text{ni}}_{d_k[c]}$$

as treatment non-inferiority where $\mathbb{I}^{\text{ni}}_{d_k[c]}$ is a boolean indicator of non-inferiority, for the comparison indexed by c within the d_k domain relative to a reference level $n_{d_k[c]}$ at a probability greater than or equal to $\gamma^{\text{ni}}_{d_k[c]}$.

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.

For any of the treatment effect parameters obtained from the primary analysis results we define:

$$\mathbb{I}^{\text{fut}}_{d_k[c]} = \Pr(\Delta_{d_k[c]} > f_{d_k[c]}) \leq \gamma^{\text{fut}}_{d_k[c]}$$

as treatment futility where $\mathbb{I}^{\text{fut}}_{d_k[c]}$ is a boolean indicator of futility with respect to the relevant assessement type, for the comparison indexed by c within the d_k domain relative to a reference level $f_{d_k[c]}$ at a probability greater than or equal to $\gamma^{\text{fut}}_{d_k[c]}$.

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 Simulation results

7 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

Chechile, R., 2020. A bayesian analysis for the mann-whitney statistic. Communications in Statistics 49, 670–696. https://doi.org/10.1080/03610926.2018.1549247

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.

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

Senn, S., 2011. U is for unease. American Statistical Association 3. https://doi.org/10.1198/sbr.2010.10024