

# CALIPSO STATISTICAL ANALYSIS PLAN

## Multicentre, adaptive, double-blind, three-arm, placebo-controlled, non-inferiority trial examining antimicrobial prophylaxis duration in cardiac surgery (CALIPSO)

### Version control

| Version name | Date       | Description / Summary of changes |
|--------------|------------|----------------------------------|
| 1.0          | 10/09/2026 | Initial version.                 |

### Associated / supporting documentation.

This statistical analysis plan (SAP) aligns with – and should be considered an accompaniment to – the CALIPSO trial protocol Version 6.0 (9 October 2024) and the ClinicalTrials.gov identifier NCT05447559.

Lead Investigator: Prof Trisha Peel, Department of Infectious Diseases, School of Translational Medicine, Monash University.

Trial Statistician: Prof Andrew Forbes, Biostatistics Unit, School of Public Health and Preventive Medicine, Monash University

This document, and any future updates, are available at <https://hdl.handle.net/11055/1463>.  
The DOI is [doi.org/10.60115/11055/1463](https://doi.org/10.60115/11055/1463).

## Table of Contents

|                                                                       |    |
|-----------------------------------------------------------------------|----|
| 1.0 Introduction .....                                                | 3  |
| 1.1 Background and Rationale.....                                     | 3  |
| 1.2 Objectives .....                                                  | 3  |
| 2.0 Trial Structure .....                                             | 3  |
| 2.1 Trial Design .....                                                | 3  |
| 2.2 Study Population.....                                             | 3  |
| 2.3 Study Interventions.....                                          | 4  |
| 2.4 Endpoints.....                                                    | 4  |
| 2.5 Randomisation.....                                                | 5  |
| 2.6 Sample size justification .....                                   | 5  |
| 2.6.1 Defining the non-inferiority margin .....                       | 5  |
| 2.6.2 Sample size calculations.....                                   | 6  |
| 2.7 Framework.....                                                    | 7  |
| 2.8 Interim analyses and stopping guidance .....                      | 7  |
| 3.0 Analysis Populations and Estimands .....                          | 8  |
| 3.1 Analysis populations .....                                        | 8  |
| 3.1.1 Primary analysis population .....                               | 8  |
| 3.1.2 Secondary analysis population.....                              | 8  |
| 3.2. Estimand for primary endpoint- SSI at 90 days post-surgery ..... | 8  |
| 3.3 Estimands for secondary endpoints.....                            | 10 |
| 4.0 Statistical analyses .....                                        | 10 |
| 4.1 Principal analyses .....                                          | 10 |
| 4.1.1 Primary endpoint.....                                           | 10 |
| 4.1.2 Secondary and safety endpoints .....                            | 10 |
| 4.2 Subgroup analyses.....                                            | 10 |
| 4.3 Missing data .....                                                | 11 |
| 4.4 Per-protocol analyses.....                                        | 11 |
| 4.5 Supplementary/sensitivity analyses .....                          | 11 |
| 4.6 Software and code fragments .....                                 | 11 |
| 5.0 References.....                                                   | 12 |

## 1.0 Introduction

### 1.1 Background and Rationale

Administration of antibiotics before incision ("surgical antimicrobial prophylaxis") is a critical infection prevention strategy in cardiac surgery. Extending doses of prophylaxis in the postoperative period is a common practice in cardiac surgery, however the benefit has not been clearly established and may drive emergence of antimicrobial resistance and lead to patient harm. The CALIPSO trial is a large international multicentre, adaptive, pragmatic, double-blind, three-arm, placebo-controlled, randomised, non-inferiority clinical trial comparing the incidence of surgical site infection after three different durations of postoperative surgical antimicrobial prophylaxis in patients undergoing cardiac surgery.

The CALIPSO trial is an adaptive, noninferiority multi-arm-multi-stage trial comparing intraoperative only (Arm A), to 24 hours (Arm B) and, to 48 hours (Arm C) of IV cefazolin and placebo postoperative surgical antimicrobial prophylaxis in 9180 patients undergoing cardiac surgery.

### 1.2 Objectives

The trial aims to assess whether:

1. Shorter prophylaxis durations are associated with a clinically acceptable low risk of surgical site infections (SSI), specifically:
  - a. Intraoperatively only is noninferior to intraoperatively plus 24 hours postoperatively
  - b. Intraoperatively only is noninferior to intraoperatively plus 48 hours postoperatively
  - c. 24-hours postoperatively differs from 48-hours postoperatively
2. The incidence of other healthcare associated infections (HCAI), including *C. difficile* are impacted by prophylaxis duration
3. The relative cost-effectiveness of the different prophylaxis durations.

## 2.0 Trial Structure

### 2.1 Trial Design

International, multicentre, adaptive multi-arm-multi-stage, pragmatic, double-blind, three-arm, placebo-controlled, phase IV, non-inferiority trial will examine the incidence proportion of SSI following cardiac surgery.

### 2.2 Study Population

Eligible patients are aged  $\geq 18$  years at the time of surgery and undergoing cardiac surgery involving a median sternotomy. Exclusion criteria are as follows:

- Age  $< 18$  years
- American Society of Anesthesiologists (ASA) physical status 5
- Patients with GFR  $< 40 \text{ mL/min/1.73m}^2$  or those requiring continuous renal replacement therapy, haemodialysis or peritoneal dialysis

- Suspected or proven surgical site infection (SSI) at previous median sternotomy site
- Patients currently receiving antibacterial (antibiotic) therapy preoperatively for treatment of a bacterial infection with a planned duration extending into the postoperative period
- Documented cefazolin hypersensitivity or severe immediate or severe delayed hypersensitivity to beta-lactams antimicrobials
- Surgery for suspected or proven endocarditis
- Surgery for cardiac transplantation
- Procedures involving insertion of a long-term (durable) ventricular assist device or mechanical circulatory support device
- Procedures not involving a median sternotomy
- Patients previously enrolled and randomised to the CALIPSO trial

## 2.3 Study Interventions

Following informed consent and eligibility screening, all participants will receive standard of care for intraoperative antimicrobial prophylaxis in accordance with the Australian Therapeutic Guidelines: Antibiotics [1], local or other national protocols or at the discretion of the treating clinician. The Australian Therapeutic Guidelines: Antibiotics recommends administration of intravenous cefazolin preoperatively within 60 minutes prior to skin incision (administration designated as ‘Time 0’), with intraoperative redosing every 3-4 hours if the procedure is prolonged [1, 2]. Our study examines the doses of prophylaxis administered *following* completion of the operation (i.e. postoperative doses). The post-operative doses of study drug ideally will commence 8 hours from time zero (i.e. after the initial dose of standard of care antimicrobial prophylaxis administered at incision) and at the discretion of the treating clinician. Participants will be randomised to receive one of the three following intervention arms (Figure 1):

**Arm A:** Placebo administered every 8-hours following the preoperative dose (time=0) for a total of five postoperative doses.

**Arm B:** Cefazolin administered 8-hourly following the preoperative dose (time=0) for two doses then placebo 8-hourly for three doses (total of 5 postoperative doses of cefazolin/placebo).

**Arm C:** Cefazolin administered every 8-hours following the preoperative dose (time=0) for a total of five postoperative doses.

## 2.4 Endpoints

**Primary** – The primary endpoint for this trial is a composite endpoint comprised of the incidence of all SSI (superficial incisional SSI, deep and, organ/space SSI) at Day 90 from index surgery (see Protocol for full details of SSI definitions).

**Secondary** – The secondary endpoints of this trial are (for definitions see Protocol):

- Incidence of *C. difficile* infection at Day 30 from index surgery
- Incidence of composite of all other HCAs (pneumonia, blood stream infection and urinary tract infection) at index discharge, at death from any cause or at day 30 from index surgery – whichever came first.
- Incidence of late SSI occurring between Day 91 and Day 180.
- All-cause mortality at 90 days from index surgery
- The individual components of the primary SSI endpoint at Day 90, namely: superficial SSI at the primary incision site, superficial SSI at the secondary incision site, deep SSI at the primary incision site, deep SSI at the secondary incision site, organ/space SSI.
- The individual components of the composite of all other HCAs at Day 30: Pneumonia,

Blood stream infections, Urinary Tract Infections and Infection-related Ventilator-Associated Complications.

Economic endpoints (a separate health economic analysis plan will be developed)

- Days alive and at home at 180 days (DAH180) after index surgery
- Direct health care costs at 180 days after index surgery in those participants meeting the primary (SSI) or secondary (HCAI) endpoints (Australian Sites and New Zealand Sites only)
- Quality of life data (EQ-5D-5L) at 30, 90 and 180 days after surgery

Safety endpoints are the incidence of adverse events including:

- Antimicrobial hypersensitivity reactions at index discharge, at time of death from any cause or at day 30 from index surgery – whichever came first.
- All-cause mortality at 180 days from index surgery
- Surgical site infections due to drug-resistant infections (defined as resistance to cefazolin) at day 90 from index surgery
- Acute kidney injury at index discharge, at death from any cause or at day 30 from index surgery – whichever came first.

NOTE: The definition the primary SSI endpoint is with reference to occurrence up to 90 days, or until death if death occurred prior to 90 days: that is, “infection while alive up to 90 days” (see Section 3.2 Estimands). In particular, patients with no infection but died prior to 90 days will be coded as “no infection”. Analogous “while alive” definitions apply to all other secondary and safety endpoints.

## 2.5 Randomisation

Randomisation will be performed at the individual patient level. Participants will be randomly assigned in a ratio of 1:1:1 to the study arms in randomly sized blocks of a multiple of 3.

Randomisation of participants to treatment arms will be as a password-protected, secure website using a central, computer-based randomisation program with permuted blocks and allocation of participants stratified according to centre and diabetes status. The treating clinicians and study investigators will have no role in the assignment process. Participants, treating clinicians, all members of the research team including the study statistician will be blinded to treatment arm allocation. In the setting of a medical emergency, such as an adverse drug reaction the code may be broken

## 2.6 Sample size justification

### 2.6.1 Defining the non-inferiority margin

The FDA and EMA have published guidance outlining approaches to determine the non-inferiority margin for non-inferiority trials [3, 4]. There are two approaches for determining the non-inferiority margin: the Fixed Margin Approach (denoted as M1) and the Synthesis Method (denoted as M2).

The Fixed Margin Approach (M1) estimates the active control effect based on analysis of past placebo-controlled trials). In this approach, the bound of the relevant 95% CI of the meta-analysis is typically selected as M1. The Synthesis Method (M2) is based on clinical judgement, that is, the largest clinically acceptable difference.

We undertook a systematic review and meta-analysis of randomised controlled trials comparing active antimicrobial prophylaxis to placebo in patients undergoing cardiac surgery. Four trials were identified [5-8]. The antimicrobials used in these trials were narrow-spectrum beta-lactams with similar activity and spectrum as cefazolin. Comparing placebo to active antimicrobial prophylaxis, the pooled absolute risk difference for SSI was 15.8% (placebo 20.2% versus active prophylaxis 3.8%: 95% CI 8.1 – 23.5). This difference (8.1%) is higher than the anticipated effect size and is clinically unacceptable. We therefore adopted the Synthesis Method, deriving the largest clinically acceptable difference (M2) as 2.5% through the Delphic principle of agreement amongst clinical experts (also less than a third of the difference of M1).

### 2.6.2 Sample size calculations

For this noninferiority trial, there are three comparisons: (1) Arm A versus Arm B, (2) Arm A versus Arm C and, (3) Arm B versus Arm C. Comparisons (1) and (2) are non-inferiority comparisons, and (3) is a superiority comparison (in either direction). The overall 5% significance level for the trial is partitioned into 1.67% for each of the 3 comparisons (1)-(3). The sample size to detect relevant differences (whether non-inferiority, harm or superiority) is based on numerical simulation of the primary endpoint (SSI), details are below. With 9180 total patients allocated to each of three, equally sized treatment arms of 3060 patients there is at least 80% power to detect a range of effects and allowing a 2% loss to follow up..

The sample size calculation is based on:

1. A 'base' percentage of SSI of 8%, obtained from the incidence of SSI from prior randomised trials which recruited from the same population and applied the data collection methods and trial definitions and in keeping with the incidence reported by Tamayo et al [9].
2. Noninferiority limit (d) 2.5%
3. The 'stopping boundaries' underlying the adaptive trial are based on:
  - a. Interim analyses are to be performed after 3060 and 6120 patients randomised, and
  - b. The interim analyses for each of 3 comparisons use symmetric O'Brien-Fleming boundaries with a two-tailed significance level of 1.67%, thereby controlling Type I error rate at 5% across all comparisons at all time points [10]. Should an arm be terminated at the first or second interim analysis, the remaining significance level will be reallocated to the comparison of the two continuing arms in the trial.

These boundaries, together with the consequent dropping of arms, and reallocation of the remaining significance level to the two continuing arms were simulated in 50,000 trials under a range of configurations to yield the following operating characteristics of the adaptive design:

1. When the true SSI rates are 8% in each arm, the probability of correctly declaring noninferiority (power) of the intraoperative arm compared to either of 24- and 48-hour arms is 88%, and power for simultaneous noninferiority compared to both these arms was 80%. The Type I error rate for the comparison of 24 versus 48 hours was 1.71%.
2. When the true SSI rates are 8% in the 24- and 48-hour arms and 10.5% in the intraoperative arm, representing the null configuration consisting of the threshold for inferiority (i.e. harm) of the intraoperative arm and equality of the 24 and 48-hour arms, the probability of falsely declaring noninferiority or harm of the intraoperative to either of 24- or 48-hours arms were 1.68% and 1.64%, respectively, and the Type I error rate for the comparison of 24 and 48 hours was 1.70%. These are all within sampling error

of 1.67%, and yield an overall familywise Type I error rate of 4.42% across all 3 comparisons and 3 analysis time points.

3. When the true SSI rates are 8% in the 24- and 48-hour arms and 13% in the intraoperative arm, representing clear harm, the probability of (correctly) declaring harm of the intraoperative arm (against either the 24- or 48-hour arms) was 89%. The Type I error for 24 versus 48 hours was 3.67% (<5%), reflecting the reallocation of alpha levels after terminating the intraoperative arm.

The familywise Type I error rate is maintained at <5% in each of the above three configurations, indicating strong control of it.

4. When the SSI rates are 8% in the 24-hour and intraoperative arms, and (albeit unlikely) 13% in the 48-hour arm, the probability of dropping the 48-hour arm at an interim analysis was 98%, and the probability of (correctly) declaring noninferiority of the intraoperative arm compared to 24 hours arm was 94%.

These results indicate that the adaptive design enables high power to detect non-inferiority when it truly exists, and proper control of Type I error rates, even when dropping of inferior arms.

## 2.7 Framework

This trial has a combination of non-inferiority and superiority hypotheses being tested. The intraoperative arm (Arm A) is to be assessed whether it is non-inferior to either or both the 24 hours arm (Arm B) or the 48 hour arm (Arm C). The comparison of 24 versus 48 hour arms is a superiority comparison.

## 2.8 Interim analyses and stopping guidance

The adaptive design allows the potential dropping of any of the three arms if clear inferiority is indicated at any of the scheduled interim analyses. These analyses will be limited to consideration of the primary endpoint (incidence of all SSI). Dropping of the intraoperative *only* arm (Arm A) may be considered if an interim analysis demonstrates that Arm A is inferior to either 24-hours (Arm B) or 48-hours (Arm C) postoperative prophylaxis. Such inferiority would require the lower endpoint of the relevant confidence interval (adjusted for interim analyses) for the difference in SSI rates (intraoperative minus 24- or 48-hours) to lie above the noninferiority margin of 2.5%. The trial would then continue randomising to 24- (Arm B) or 48-hours (Arm C) to provide the definitive answer. Similarly, if either the 24-hours (Arm B) or 48-hours (Arm C) demonstrates conclusive evidence of a true difference in their head-to-head comparison at interim analysis, via the confidence interval for the difference excluding the zero point, dropping the inferior of the two arms may be considered, and the trial continued with just the superior arm and the intraoperative arm (subject to it too not demonstrating inferiority). If the interim analyses shows intraoperative only (Arm A) is noninferior to one or both of 24- (Arm B) and 48-hours (Arm C), as evidenced by the upper endpoint of the confidence interval for the difference in SSI rates lying below 2.5%, the trial **will not** terminate, as there is no ethical dilemma in continuing, and given the importance of determining with maximum certainty of the optimal duration of prophylaxis to assist with translation of these currently embedded practices. Early stopping of any arm may also

be considered by the DSMC based on designated safety endpoints. Stopping the trial for futility will not be performed.

As described in the sample size section (2.6.2), the interim analysis stopping boundaries are based on the approach of Follman et al [10], in which each of the 3 pairwise comparisons between treatment arms is performed at an overall  $0.05/3 = 0.0167$  level (Bonferroni), with critical values for each comparison determined using a symmetric O'Brien-Fleming alpha spending function, and truncated at a maximum of  $\pm 4$ . The exact boundaries at each interim analysis will be determined based on the information fraction attained at that actual analysis time (i.e., proportion of total number of patients reaching the 90 day followup point), using the `ldBounds()` function in the R `ldbounds` package. For example, if the interim analyses are performed at information fractions of 1/3, 2/3 and 1, the boundary values (Z-statistics) are  $\pm 4$ ,  $\pm 3.03$ ,  $\pm 2.41$  at the 1<sup>st</sup> interim analyses, 2<sup>nd</sup> interim analysis and final analysis, respectively, corresponding to nominal two-tailed (superiority) p-values of 0.00006, 0.0024, and 0.016, and confidence intervals with levels 99.994%, 99.76%, 98.40%, respectively.

## 3.0 Analysis Populations and Estimands

### 3.1 Analysis populations

#### 3.1.1 Primary analysis population

The primary analysis population consists of all patients randomised to a treatment arm who undergo cardiac surgery involving a median sternotomy and are analysed according to the regimen to which they were initially allocated irrespective of any deviations from this regimen or any other protocol deviations. We follow the principles outlined in Fergusson et al. [11] to determine any post-randomisation exclusions, such as excluding patients who withdrew consent or whose clinician withdrew the patient from the trial and for whom no post-randomisation data was collected; or who did not receive their study treatment because it was not available at their site. Patients discovered to be ineligible after randomisation by default will be included in this analysis population unless, based solely on information obtained pre-randomisation, the reason for (and discovery of) ineligibility is determined to be unrelated to treatment arm and factors prognostic of outcome, in which case they will be excluded. The resulting population after any such exclusions is commonly known as the modified intention-to-treat population (mITT). This primary analysis population will inform the primary estimand (see Section 3.3).

#### 3.1.2 Secondary analysis population

Per protocol analyses may also be performed in a secondary analysis population that is restricted to patients who completed the treatment to which they were allocated, meaning only those who receive all doses of cefazolin/placebo according to their original allocation. This is often called the “per protocol (PP)” population.

### 3.2. Estimand for primary endpoint- SSI at 90 days post-surgery

CALIPSO will incorporate the estimand framework [12] in the analysis of primary and secondary endpoints. An estimand is a precise description of the treatment effect reflecting the clinical question posed by a given clinical trial objective. It has five attributes: population, treatment,

endpoint of interest, summary measure, and possible intercurrent events (defined as an event that can occur post-randomisation and preclude or affect the interpretation of the endpoint). Note that unless otherwise specified, all endpoints are measured from Day 0 (day and time of surgery).

Objective: To estimate the effect of treatments in patients undergoing cardiac surgery on the incidence of SSI (any of superficial, deep, and organ space), as defined by the CDC criteria, up to 90 days after surgery

Calculation: SSI is recorded at hospital discharge, 30, 90 and 180 days post-surgery, and any reported possible SSI at each of these time points undergoes endpoint adjudication. The primary endpoint is a binary variable SSI to 90 days, calculated as “Yes” if (adjudication) confirmed SSI occurs at any time point up to 90 days, and “No” if no SSI is recorded at each of the time points up to 90 days. Patients with no confirmed SSI but are missing entries for possible SSI at discharge, 30 days or 90 days, will have their SSI to 90 days set to “Missing”.

The following intercurrent events are currently considered for the primary SSI endpoint:

- Cancellation or postponement of surgery
- Death prior to endpoint followup time
- Non-adherence to allocated treatment (which includes incomplete/underdosing, treatment crossover/switching, non-adherence to the assigned intervention and treatment discontinuation).
- Early return to operating theatre within 48 hours of surgery.

The 5 attributes of the estimand related to the primary endpoint are displayed in Table 1.

**Table 1.** Estimand for the CALIPSO primary endpoint

|                                                                                                                                                               |                                                                                                                |                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Target population</b>                                                                                                                                      | Patients meeting the PROMPT eligibility criteria and randomised to a treatment.                                |                                                                                                                                                                                     |
| <b>Interventions</b>                                                                                                                                          | Durations of antibiotics as defined by Arms A,B,C in section 2.3.                                              |                                                                                                                                                                                     |
| <b>Endpoint</b>                                                                                                                                               | Incidence of any of superficial, deep and organ/space surgical site infection up to Day 90 from index surgery. |                                                                                                                                                                                     |
| <b>Population-level summary measure</b>                                                                                                                       | Risk difference comparing each treatment arm to each other treatment arm                                       |                                                                                                                                                                                     |
| <b>Intercurrent events</b>                                                                                                                                    | Handling of event                                                                                              | Application to full dataset                                                                                                                                                         |
| 1. <i>Cancellation or postponement of surgery</i>                                                                                                             | <i>“principal stratification” strategy by simple exclusion.</i>                                                | Exclude patients with surgery cancelled or postponed. This is valid since ineligibility would have occurred regardless of the treatment arm allocation – “always surgery” patients* |
| 2. <i>Death prior to 90 days</i>                                                                                                                              | <i>“while alive” strategy.</i>                                                                                 | Patients dying prior to 90 days and without an SSI shall be coded as “No SSI”.                                                                                                      |
| 3. <i>Non-adherence to allocated treatment (which includes treatment crossover, non-adherence to the assigned intervention and treatment discontinuation)</i> | <i>“Treatment policy” strategy</i>                                                                             | Treatment assignment will be as randomised, regardless of non-adherence or return to theatre                                                                                        |
| 4. <i>Early return to operating theatre</i>                                                                                                                   |                                                                                                                |                                                                                                                                                                                     |

\*This is the commonly used mITT population, in which exclusion is unrelated to treatment randomised.

The primary estimand will be the incidence of the primary endpoint (SSI) (to 90 days or death) for each treatment relative to each other treatment after index surgery for the primary analysis population.

### 3.3 Estimands for secondary endpoints

The secondary endpoints in Section 2.4 will use the same estimand strategy as for the primary endpoint.

## 4.0 Statistical analyses

The analysis and reporting of the results will follow the CONSORT guidelines [13]. Data will be summarised with respect to baseline characteristics, primary, secondary and safety endpoints. Data will be presented in summary tables according to treatment allocation. As CALIPSO is a large randomised controlled trial, covariate balance is expected and no covariate adjustment will be performed apart from adjustment for the randomisation stratification factor of diabetes status.

### 4.1 Principal analyses

#### 4.1.1 Primary endpoint

The principal analysis of the binary primary endpoint will use the primary analysis population (section 3.1.1, mITT) to estimate the differences in incidences of SSI between each pair of treatment arms using a single binomial regression model with an identity link function, with fixed effect terms for the treatment groups and adjusting for diabetes status at randomisation. Confidence intervals will be adjusted for multiplicity of interim analysis assessments and for the 3 pairwise comparisons as specified in the Interim Analysis section 2.8. Specific levels of confidence intervals coverage and guidelines for interpretation of these intervals for each of the 3 pairwise treatment comparisons are also described in Section 2.8.

#### 4.1.2 Secondary and safety endpoints

All reported AEs and SAEs will be summarised in aggregate and by intervention.

Because all secondary and safety endpoints are binary, the analysis of each secondary endpoint will use the same model as for the primary endpoint, except that confidence interval levels will be 98.33% ( $=100\% - 5\%/3$ ) to control for 3 pairwise comparisons, i.e. without adjustment for interim analysis multiplicity.

### 4.2 Subgroup analyses

Subgroups analyses for the primary endpoint will be conducted for age ( $\leq 60$ ; 61–70; 71–75;  $> 75$ ), sex and body mass index ( $\leq 18.5$ ;  $> 18.5$ –25.0;  $> 25.0$ –30.0;  $> 30.0$ –35.0;  $> 35.0$ ). These analyses will be performed by adding treatment X subgroup interaction terms to the binary regression model. Pairwise treatment comparisons will be reported separately for each subgroup factor, however interaction p-values will guide the overall evidence for variation in these treatment comparisons.

### 4.3 Missing data

Missing baseline characteristics will be imputed using single mean imputation pooled across treatment arms.

A low incidence of the primary endpoint being missing due to loss to followup is expected, however if it exceeds 5%, a multiple imputation approach will be utilised. This will utilise baseline and post-baseline variables predictive of missingness (e.g with correlation with missingness of  $>0.10$ ), and that are also correlated with SSI (in those with SSI present)  $> 0.10$ , with the purpose of the latter to increase precision. Imputation will be separately by treatment group using chained equations with predictive mean matching and 5 nearest neighbours, and repeated to produce 20 imputed datasets that are then combined with differences in SSI rates between treatment groups calculated using Rubin's rules.

### 4.4 Per-protocol analyses

Per-protocol analyses of the primary endpoint will be performed in the secondary analysis (PP) population (see definition of compliance in Section 3.1.2) in order to assess the effect of randomised treatment in the hypothetical case of complete treatment adherence. This will use the same analysis methods as for the primary analysis (Section 4.1.1), but with the addition of regression adjustment for baseline and intraoperative variables predictive of protocol non-adherence (e.g., correlation  $> 0.10$ ) and prognostic of the SSI endpoint (correlation  $> 0.10$ ) in order to control for potential selection bias arising from the patient exclusions from the primary (mITT) population. This regression adjustment is valid if the adjustment variables are sufficient to make plausible the assumption that treatment non-adherence is 'at random' conditional on randomised treatment arm and these adjustment variables.

### 4.5 Supplementary/sensitivity analyses

Analyses of the primary endpoint will be repeated adjusted for study site as a random effect. If convergence is not obtained then standard errors will be adjusted using cluster-robust standard errors by site.

### 4.6 Software and code fragments

All analyses will be undertaken in Stata version 18 or later

Sample Stata code for the primary endpoint will use the `binreg` command:

```
binreg outcome i.TreatmentGroup i.Diabetes_Strata, rd
```

Adjustment for site as a random effect:

```
xi: gllamm outcome i.TreatmentGroup i.Diabetes_Strata, i(site) fam(binomial)
link(iden) adapt nip(12)
```

Standard errors adjusted for clustering by site:

```
binreg outcome i.TreatmentGroup i.Diabetes_Strata, rd cluster(site)
```

## 5.0 References

1. Antibiotic Expert Group. *Therapeutic Guidelines: Antibiotic*. 16th edn. Melbourne: Therapeutic Guidelines Limited, 2019.
2. Bratzler DW, Dellinger EP, Olsen KM, *et al*. Clinical Practice: Guidelines for Antimicrobial Prophylaxis in Surgery. *Surg Infect (Larchmt)* 2013;14:73–156.
3. European Medicines Agency (EMA). Guideline on the Choice of the Non-Inferiority Margin. In: Committee for Medicinal Products for Human Use, ed. London: EMEA; 2005.
4. Food and Drug Administration (FDA). Non-Inferiority Clinical Trials to Establish Effectiveness: Guidance for Industry. U.S. Department of Health and Human Services; 2016.
5. Fekety FR, Jr., Cluff LE, Sabiston DC, Jr., Seidl LG, Smith JW, Thoburn R. A study of antibiotic prophylaxis in cardiac surgery. *J Thorac Cardiovasc Surg* 1969;57:757-63.
6. Fong IW, Baker CB, McKee DC. The value of prophylactic antibiotics in aortic-coronary bypass operations: a double-blind randomized trial. *J Thorac Cardiovasc Surg* 1979;78:908-13.
7. Goodman JS, Schaffner W, Collins HA, Battersby EJ, Koenig MG. Infection after Cardiovascular Surgery. *New England Journal of Medicine* 1968;278:117-23.
8. Goodman JS, Schaffner W, Collins HA, Battersby EJ, Koenig MG. Infection after cardiovascular surgery. Clinical study including examination of antimicrobial prophylaxis. *N Engl J Med* 1968;278:117-23.
9. Tamayo E, Gualis J, Flórez S, *et al*. Comparative study of single-dose and 24-hour multiple-dose antibiotic prophylaxis for cardiac surgery. *J Thorac Cardiovasc Surg* 2008;136:1522–7.
10. Follmann DA, Proschan MA, Geller NL. Monitoring pairwise comparisons in multi-armed clinical trials. *Biometrics*. 1994 Jun;50(2):325-36.
11. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002; **325(7365)**:652-4.
12. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2020. [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf).
13. Hopewell S, Chan AW, Collins GS *et al*. CONSORT 2025 Statement: Updated Guideline for Reporting Randomized Trials. *JAMA*. 2025 Jun 10;333(22):1998-2005. doi: 10.1001/jama.2025.4347.