



Informal notes and discussion points

ROADMAP: RandOmised Arthroplasty infection worldWide Multidomain Adaptive Platform trial - simulation report

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

Version: 0.1

Date: 26 May 2025

Authors: Mark Jones and James Totterdell

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

Table of contents

| | | |
|-----------|---|-----------|
| 1 | Section 1.1 - Background | 4 |
| 1.1 | Discussion | 4 |
| 2 | Section 1.3 - Estimands | 5 |
| 3 | Additional/unnecessary material | 6 |
| 3.1 | Discussion | 6 |
| 3.2 | Aside | 7 |
| 3.3 | Discussion | 7 |
| 4 | Section 1.3.1.1 - Estimand B.1 (surgical intervention) | 8 |
| 4.1 | Update required | 8 |
| 4.2 | Discussion | 8 |
| 4.3 | Discussion!!!! | 8 |
| 4.4 | Discussion required | 9 |
| 4.5 | Discussion/update required | 9 |
| 4.6 | Discussion | 9 |
| 5 | Section 1.3.1.2 - Estimand D.1 (antibiotic duration intervention) | 10 |
| 5.1 | Discussion | 10 |
| 6 | Section 1.3.1.3 - Estimand E.1 (extended prophylaxis intervention) | 12 |
| 7 | Section 1.3.1.4 - Estimand C.1 (antibiotic choice intervention) | 13 |
| 8 | Section 1.3.2.1 - Estimand B.2 (surgical intervention) | 14 |
| 9 | Section 1.4.3 - Data management | 15 |
| 10 | Section 2.1.2 - Per-protocol | 16 |
| 11 | Section 2.2 - Subgroups | 17 |
| 12 | Section 2.3 - Descriptive summaries | 18 |
| 13 | Section 2.4 - Sequential analyses | 19 |
| 14 | Section 2.6 - Primary analysis | 20 |

| | | |
|----|---|----|
| 15 | Section 2.6.1 - Priors | 21 |
| 16 | Section 2.9 - Supportive analyses | 22 |
| 17 | Section 2.9.1 - DOOR | 23 |
| 18 | Section 2.9.11 - Time alive and free from revision procedure to 24 months | 24 |
| 19 | Section 2.10 - Supportive (domain specific) analyses | 25 |
| 20 | Section 2.10.1.1 - Treatment success at 12 months | 26 |
| 21 | Section 2.10.2.2 - Acute liver injury following platform entry to 90 days | 27 |
| 22 | Section 2.10.2.3 - Laboratory-proven <i>Clostridium difficile</i> diarrhoea to x days | 28 |
| 23 | Section 2.10.2.4 - Antibiotics ceased due to suspected adverse reaction to 90 days | 29 |
| 24 | Section 2.10.3.2 - Time alive and free from any revision procedure to 12-months | 30 |
| 25 | Section 2.10.3.3 - Time alive and free from any revision procedure to 24-months | 31 |
| 26 | Section 2.10.3.4 - Laboratory-proven <i>Clostridium difficile</i> diarrhoea to x days | 32 |
| 27 | Section 2.10.3.5 - Antibiotics ceased due to suspected adverse reaction to 90 days | 33 |
| 28 | Section 2.10.4.2 - Acute liver injury to day 100 | 34 |
| 29 | Section 2.10.4.3 - Acute liver injury to day 100 | 35 |
| 30 | Section 2.10.4.4 - Laboratory-proven <i>Clostridium difficile</i> diarrhoea to x days | 36 |
| 31 | Section 2.10.4.5 - Antibiotics ceased due to suspected adverse reaction to 100 days | 37 |
| 32 | Section 3.1 - Treatment effects | 38 |
| 33 | Section 3.1.1 - Surgical domain | 39 |
| 34 | Section 4.1 - Decision procedures overview | 40 |

1 Section 1.1 - Background

1.1 Discussion

I previously included a statement on how we should interpret missing elements in the composite as a failure case (or alternative approach required) but I think I just wanted to know (1) how we should be handling this and (2) how likely it is to happen.

The core protocol, see section 6.6 Trial endpoints does not seem to indicate what happens when elements of the composite cannot be identified.

2 Section 1.3 - Estimands

3 Additional/unnecessary material

Where the treatment policy strategy is used, any ICEs are effectively ignored and/or considered part of the treatment protocol. Where other methods of handling ICEs are used, it is important to recognise that treatment-specific ICEs may occur in isolation for a given unit or may occur in conjunction with ICEs applicable to treatments from other domains. For example, a patient may switch from DAIR to two-stage revision due to having a loose joint but also discontinue extended prophylaxis due to adverse reaction. In such cases, the ICEs are still handled by the applicable strategy for the estimand under consideration. For the above example, this implies 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. This perspective aims to be simple, but it might do so at the cost of being naive.

3.1 Discussion

In the core protocol section 6.7.1 it has “To evaluate, within each relevant cell or combination of cells (as per the DSA), the effect of revealed randomised interventions compared to the domain control, on the probability of treatment success at 12 months after platform entry, in platform eligible participants”

What is meant by the reference to “within each relevant cell”?

In the Surgical late acute dsa section 9.1.1 “To evaluate, within the domain, the effect of revealed randomised intervention (revision) compared to the domain control (DAIR) on the probability of treatment success at 12 months after platform entry, in domain eligible participants.”

The above is silent on the status of the other variables that contribute to the probability of treatment success, I would interpret the underlying goal as one of trying to obtain a marginal effect (averaged over the sample).

3.2 Aside

Conditional OR: Among two hypothetical individuals with identical covariates $X = x$, one given $A = 1$ and one $A = 0$, the odds of death differ by a factor of $\exp(\beta_A)$. Preference when making individual level inference or when covariates are effect modifiers.

Marginal OR: If the entire study population were switched from $A = 1$ to $A = 0$, the overall odds of death would change by OR_{marg} . More relevant to a policy context or when comparing to an unadjusted analysis.

3.3 Discussion

The estimands are introduced in generality in the core protocol and made specific in the DSA. I feel like the estimands are not currently providing much value because there isn't enough detail and they do not really seem to consider the implications of ICEs on the existence or interpretation of the outcome. The purpose of an estimand specification is to ensure that the trial answers a very precise clinical question in a way that is clear and interpretable to all stakeholders (e.g., clinicians, regulators, patients), aligns with the trial objectives, and is both statistically valid and estimable from the collected data. If the estimands are inadequately specified, this could lead to criticism.

Given that the majority of the analyses follow a treatment-policy, I am bit torn as to whether the spending the time on clarifying things is going to make any real difference to anyone. One thing we could do to make things easier on us is to minimise duplicating the estimands across documents. For example, we might remove the estimands in the core protocol, which are not really adding anything, and just retain the DSA definitions. The rationale for this stems from the fact that the core protocol specifications do not relate to any single intervention and therefore do not really address the issues that estimands are intended to address, i.e. they do not provide clarity on what effects we are targeting nor on the ICEs as noted above. As a side note, in my opinion, I feel that if the overall objective is on assessing risk, then that should be what we are aiming to infer analytically rather than odds ratios and that is what is presently adopted in the sap.

4 Section 1.3.1.1 - Estimand B.1 (surgical intervention)

4.1 Update required

On (Page 19, ROADMAP Surgical Late Acute DSA. V1.1_01Aug2024.pdf) there is a note in parentheses that states “this will be estimated within hip versus knee strata”. If the intention is to have joint specific focus on the surgical treatment effects then this should be explicitly stated as part of the estimand objective.

At the stats meeting dated 9/12/2024, we decided to apply stopping rules to the aggregated hip and knee groups rather than adapting the trial based on joint specific effect. However, both overall and joint specific effects are to be reported (see later). This is not yet reflected in a released protocol.

4.2 Discussion

If a dair switches to one-stage (or two-stage) does this patient then enter into the respective duration domain?

4.3 Discussion!!!!

Currently, we obtain the effect of revision under the notion of apples versus apples. The only context in which this arises is when d2 and d3 are set to the non-randomised level.

An alternative practice is to contrast dair (plus d2 and d3 as given as usual practice under dair, which happens to be the non-randomised set to these domains) versus revision (plus d2 and d3 as given per the usual practice under revision and specifically the type of revision adopted). I think the latter is what is actually required (and makes most sense, at least to me at the moment).

4.4 Discussion required

Other ICEs are possibly relevant but not yet considered. Withdrawal based on clinician opinion on patient's interests with regards to their involvement in roadmap. 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 as variations in the accepted procedures (i.e. different from the protocol definitions) but not leading to switching also constitute an ICE that could impact the interpretation of the treatment effects. An example could be secondary surgery that does not lead to replacement of prosthetic, i.e. not handled via the composite strategy. Deviations from any protocol timings such as a requirement to have a surgical procedure within a given time window from randomisation. Also, it should be considered whether switchover should be interpreted as a failure rather than an ICE. Currently, any additional unspecified ICE would also be handled via the treatment policy strategy

4.5 Discussion/update required

I feel like risk differences are maybe preferable as they are more directly interpretable than odds ratios, albeit a bit more work to the analyst. However, the implication is that we need to be explicit about which population we are marginalising over but all that entails is saying that we are averaging over the sample. Typically, this would be the patients in the trial but given our comparisons are very specific, the levels which we use in the g-computation are also specific for some terms. The random effects will probably be taken to be zeros and so our inference will be interpretable for the 'typical' site (or whatever the random effect pertains to).

4.6 Discussion

I am wondering what the implications of switchover are on entry into the other domains. For example, a patient switches from DAIR to 2-stage revision. In theory, after the switch, they are eligible for entry into extended prophylaxis, but randomisation has already decided that they will receive non-randomised treatment for this domain.

Is this a problem? What should be done here?

5 Section 1.3.1.2 - Estimand D.1 (antibiotic duration intervention)

5.1 Discussion

The DSA indicates “Participants will be randomised at platform entry in a fixed 1:1 ratio across the domain.” where “platform entry” is defined as “the timepoint when the patient has met core eligibility criteria, given informed consent for the platform, and been randomised” in the core protocol.

Get confirmation that **receipt** of one-stage revision is the correct population definition.

For all silos, the decision on the classification of surgery type appears to be made on completion of the surgery when the surgeon indicates that there are no plans for a second stage.

In the DSA for abx duration it states: “Platform eligible participants within any of the silos who have had a single stage revision will be randomised to receive either 6 weeks (short-course antibiotics) or 12 weeks (long-course antibiotics).”

Does that mean that someone who was in the late-acute silo and that was randomised to DAIR, but switched to one-stage revision due to a clinical decision (perhaps because of a loose joint) will then enter into the abx duration domain?

This setup suggests that allocations are generated across all domains for all patients irrespective of whether their dependencies negate the possibility of entry? For example, someone in late-acute is allocated to DAIR, but nevertheless they are also allocated to both the duration domains and the choice domain. I believe this is the case, but very much need to confirm.

So, patients can only be revealed to randomised assignment in the abx duration domain if they survive surgery. Implies start of follow up and treatment initiation not coincident, implies analysis is actually a landmark analysis and should be reflected as such for the relevant domains.

If a patients dies during surgery, what is their treatment designation?

If they do die (assuming we can obtain a treatment designation as one-stage), they are failure cases and would not be revealed to either of the abx duration. Isn't this exactly the setup in which immortal time bias occurs?

We are not evaluating duration effects in the population, we are looking at duration effects based on a post-randomisation event, i.e. survival through surgery. We are selecting based on survival. Yes, but the analysis is landmarked and we still have a randomised comparison.

6 Section 1.3.1.3 - Estimand E.1 (extended prophylaxis intervention)

ICE and treatment regimen:

todo - Get more detail on E.1(i2) and check my interpretation is correct. Would the ineligibility actually mean that they do not receive the extended prophylaxis but would receive extended backbone antibiotic?

7 Section 1.3.1.4 - Estimand C.1 (antibiotic choice intervention)

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.

8 Section 1.3.2.1 - Estimand B.2 (surgical intervention)

Should the secondary estimands be removed?

High level lay overview:

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 most well aligned approach. But this is very hard to implement in our setting. Are there any other options?

ICEs and treatment regimen:

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 when it is not possible to remove the existing implant. 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.

9 Section 1.4.3 - Data management

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

10 Section 2.1.2 - Per-protocol

Principal stratum is referred to as a per-protocol in the DSAs but this is incorrect.

Per-protocol analyses traditionally exclude those patients that deviate from the protocol and conduct the primary analysis with this subset. 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.

Specific questions on the protocol definitions:

- For the 6wk and 12wk antibiotic duration assignment, is there a reqt for no abx after 16 wks?
- For the extended prophylaxis, is there a reqt for no abx after 90 days?
- For the Rifampicin, is there any requirement for the interval between the doses?
- Are we only considering per-protocol with respect to treatment?

11 Section 2.2 - Subgroups

Title should be updated to subgroup analyses, not secondary analyses for:

DSA Surgical Late-acute section 9.6 DSA for AB duration part A section 9.6 DSA AB Duration part B section 8.6 (toc does not link correctly) DSA AB Choice section 9.6 (this has both secondary and subgroup analyses, keeping a clear separation between the two would be helpful)

Clarify how duration of symptoms is intended to be categorised for surgical domain.

For surgical domain, categorise time from implantation or get rid of.

For backbone duration, discuss removing ideal procedure vs not as this suggests a protocol deviation.

For ext proph, categorise duration between first and second and duration of antibiotic treatment between first stage and reimplant.

For ab choice, include silo or was that purposefully excluded as rif is expected to work by the same action?

There was a review question at the top of the subgroup section for which clarification is needed - *is it obvious what this means (in the parameter estimation)?*

For one-stage vs two-stage - this is a post-randomisation subgroup, so could be influenced by previously revealed assignments, eg rifampicin. AB Choice is revealed and administration commenced following the first procedure, i.e. after a DAIR, one-stage or between the first and second stage of a two-stage revision and therefore, the reveal of AB choice will not influence selection of surgery type.

12 Section 2.3 - Descriptive summaries

Query how the consent process works for say AB choice where eligibility is not assessed until after the initial surgical treatment?

13 Section 2.4 - Sequential analyses

Revisit the timing of interim analyses as enrolment always seems to be a major challenge.

I think it is probably more operationally practical to run interims on a regular timing rather than for fixed sample size, but I still generally prefer the latter for simplicity of understanding in the design operating characteristics.

The subgroup analysis at each interim used to determine whether the trial should continue is problematic because

1. it opens us up to severe criticism of un-specified decisions occurring
2. type I inflation

14 Section 2.6 - Primary analysis

Incorporate silos as random effect on d1 treatment deviations?

Revisit the baseline characteristics to be included in the model. Reference some of the sub-groups for ideas.

15 Section 2.6.1 - Priors

todo - revisit baseline covariates with clinical group.

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

16 Section 2.9 - Supportive analyses

TODO - Query the competing risks, what is relevant/sensible to consider?

Should we revise the estimand definition such that we state clearly the issue of competing risk and that it is a part of the estimand that is being targeted (or change the estimand and estimation approach completely)?

17 Section 2.9.1 - DOOR

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, which can generally only be obtained from a cross-over trial. For example, a PI of 0.7 means that if you randomly select a patient from the experimental group and another from the control group, there is a 70% chance the experimental group patient will have a better outcome. This is not the same as saying that 70% of patients will benefit from the experimental treatment. For further cautions, limitations and critique on interpretation see Senn ([2011](#)).

A Bayesian approach to the MWU does exist, but has not been widely adopted to date and so we have opted for the above approach. Chechile ([2020](#)) provides details on the Bayesian perspective.

18 Section 2.9.11 - Time alive and free from revision procedure to 24 months

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, fit for purpose and so on.

todo - This deviates from the current estimand specification A.12 (Pg 39 ROADMAP Core Protocol_V1.1_01Aug2024_clean.pdf) which, I believe, intends a descriptive summary.

19 Section 2.10 - Supportive (domain specific) analyses

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

20 Section 2.10.1.1 - Treatment success at 12 months

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.

21 Section 2.10.2.2 - Acute liver injury following platform entry to 90 days

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.

22 Section 2.10.2.3 - Laboratory-proven Clostridium difficile diarrhoea to x days

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

23 Section 2.10.2.4 - Antibiotics ceased due to suspected adverse reaction to 90 days

todo - What is meant by other?

Check whether this was meant to be 90 days post commencement of antibiotics.

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

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

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

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

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

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

27 Section 2.10.3.5 - Antibiotics ceased due to suspected adverse reaction to 90 days

todo - What is meant by other?

Check whether this was meant to be 90 days post commencement of extended prophylaxis.

28 Section 2.10.4.2 - Acute liver injury to day 100

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

29 Section 2.10.4.3 - Acute liver injury to day 100

todo - Check if c.4 is a duplicate??

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

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

31 Section 2.10.4.5 - Antibiotics ceased due to suspected adverse reaction to 100 days

todo - Check why this is 100 days and others are 90. Should these all be made the same?

32 Section 3.1 - Treatment effects

Old notes - review and recast if necessary:

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 $E[Y|T = 1] = y_1$ and $E[Y|T = 2] = y_2$ (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, $E[Y(1)|T = 1] \neq E[Y(1)|T = 2] \neq E[Y(1)]$ and $E[Y(2)|T = 1] \neq E[Y(2)|T = 2] \neq E[Y(2)]$. 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.

33 Section 3.1.1 - Surgical domain

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

34 Section 4.1 - Decision procedures overview

Define what happens to subsequent participants within a domain for which the decision has been dealt with. Presumably, the simplest approach is to simply add them into a non-randomised comparison for the domain for which the decision was made, but to introduce them into other domains based on the design dependencies.

Unsure whether it is sensible to be examining all subgroups in order to determine the outcome of futility, especially given the number of subgroups defined.

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>

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