Master Statistical Analysis Plan for the PRACTICAL Platform

Prepared using Gamble et al. (2017) Guidelines for the Content of Statistical Analysis Plans in Clinical Trials, JAMA. 2017;318(23):2337-2343.

Title and Trial Registration
Platform of Randomized Adaptive Clinical Trials in Critical Illness (PRACTICAL)
Randomized Controlled Trial
Trial registration number: NCT05440851

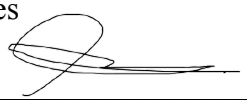
SAP Version
Master Statistical Analysis Plan Version 2.0
Date: 30-Aug-2022

Protocol Version
Master Protocol Number OZUHN-004
Version 3.0

SAP Revisions
Original: Version 1.0 dated 03Dec2021

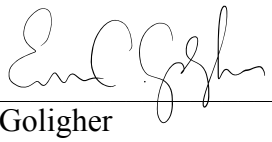
SAP Contributors
Anna Heath
Christopher Yarnell
Wei Xu
Martin Urner
Ewan Goligher
Eddy Fan
Niall D Ferguson
George Tomlinson

Signatures



September 12, 2022

Anna Heath Date
Chair of Statistical Analysis Committee



September 12, 2022

Ewan Goligher Date
Platform Sponsor

Contents

Section 1: Introduction ... 3
Background and rationale ... 3
Objectives ... 3
Section 2: Study Methods ... 3
Trial design ... 3
Randomization ... 4
Sample size ... 5
Framework ... 5
Statistical interim analyses and stopping guidance ... 5
Timing of final analysis ... 6
Timing of outcome assessments ... 7
Section 3: Statistical Principles ... 7
Statistical Reporting ... 7
Adherence and protocol deviations ... 7
Analysis populations ... 7
Section 4: Trial Population ... 8
Screening data ... 8
Eligibility ... 8
Recruitment ... 9
Withdrawal/follow-up ... 9
Baseline patient characteristics ... 9
Section 5: Analysis ... 9
Outcome definitions ... 9
Analysis methods ... 11
Missing data ... 12
Additional analyses ... 13
Harms ... 13
Statistical software ... 13
Bibliography ... 14
Appendix: Inclusions for Domain-Specific Protocols/Analysis Plans ... 15

Section 1: Introduction

Background and rationale

Acute hypoxemic respiratory failure (AHRF) is a common, life-threatening condition associated with substantial morbidity and mortality. Many acute and critical illnesses can cause AHRF, defined as a requirement for invasive or non-invasive ventilatory support with a PaO2/FiO2 ratio of 300 mm Hg or less. Patients with AHRF are vulnerable to lung and diaphragm injury associated with mechanical ventilation along with other nosocomial complications of critical care. While substantial progress has been made to reduce these risks and improve outcomes, many aspects of care for patients with AHRF remain untested and poorly defined. Novel approaches require careful evaluation in early phase studies to assess their potential to benefit patients before definitive evaluation in phase III trials. This platform trial will enable pilot/feasibility, phase II and phase III evaluation of a range of potential treatments to improve outcomes for patients with AHRF.

Objectives

The overarching objective of the PRACTICAL platform trial is to facilitate the development and evaluation of novel interventions for patients with AHRF. The primary endpoint for evaluation of these interventions will be specified for each domain, depending on its objectives and phase of investigation. This document is the Master SAP outlining the principles of randomization and analysis for this trial.

Section 2: Study Methods

Trial design

This study will use a randomised, embedded, multifactorial adaptive platform (REMAP) design [1]. This design is the basis of a perpetual trial to assess novel therapies for patients with AHRF efficiently and effectively. The trial is structured across independent *domains*, $d = A, B, C \ldots$, where patients are allocated to two or more interventional arms, $j = 1, \ldots, J_d$ (where J_d represents the total number of treatments in domain d). We will denote treatment j in domain d by d_j.

Domains and interventions added into a currently enrolling domain may first employ a pilot or feasibility phase. Interventions on the platform studied at the pilot or feasibility phases will not employ interim statistical monitoring. Interventions studied at the phase II stage will focus on shorter term/surrogate endpoints with a goal of graduating to a phase III evaluation on the domain-level primary outcome if potential benefit is seen in the phase II evaluation. When appropriate, domain-specific statistical analysis plans will specify rules for seamless graduation from phase II to phase III. Domains will use a common phase III endpoint but can consider different phase II endpoints depending on the intervention.

Domain-specific eligibility will be noted in domain specific protocols and may be dependent on the treatment that the individual is randomised in another domain. The REMAP design aims to offer individuals treatments from multiple domains. If interventions are given concurrently, interaction effects may be considered and will be pre-specified in the domain specific protocols and SAPs. To avoid potential deleterious clinical interactions, enrolment into a given domain may be specified to preclude enrolment into other domains where appropriate.

Domain-specific eligibility will be dependent on a patient's *state,* $s = 1, \ldots, S$. States are a set of S mutually exclusive clinical conditions, defined by a patient's disease characteristics. The patient's state can change over the course of their enrolment in the trial, implying that a patient's eligibility for different domains could change. Treatment effects are expected to differ across the states. Thus, the aim of this trial is to estimate state-specific treatment effects within each domain.

Randomization

Randomization for a specific domain will occur when a patient is enrolled and consented into the domain. Randomization in a specific domain will be independent of previous treatment assignment, except in cases where the randomization is dependent on other domains, as specified in the domain specific protocol. We will use the word *regimen* to describe the complete set of treatments that an individual is randomised to, across each domain in which they were randomized. Within each domain, we will use block randomisation with random block sizes. Block sizes will change depending on the domain as they will include different numbers of arms. If applicable, domains will start with equal randomisation for the pilot/feasibility phase. In some domains, we will consider response adaptive randomisation after the feasibility phase, which we outline below. Patients will be randomised within a secure, online, centralized randomization tool housed at Ozmosis Research.

Stratification

In general, randomization will be stratified by site and, for domains that enrol patients across multiple states, by state. Randomization in a specific domain could also be stratified based on additional factors, as identified in the domain-specific protocol. These factors may be related to the patient but should be *stable* across time, i.e., age, sex. Due to the set-up of the randomization tool, all stratification factors will be determined at the start of the domain and will not change. Characteristics of the patient's disease that could change across the treatment period will designate the patient's *state* and alter their eligibility for domains.

Response Adaptive Randomisation

We may consider using response adaptive randomisation (RAR) for some domains, based on ethical, clinical, and statistical considerations. For statistical considerations, simulations will explore whether RAR will result in meaningfully greater efficiency and/or an expected reduction in patient harm.

RAR will initialise by enrolling a domain-specific number of participants using equal randomisation, known as the "burn-in" period for the RAR. This burn-in period will usually include all the participants in the pilot/feasibility phase, if applicable. After this burn-in period, the proportion of participants randomised to treatment will be related to the posterior probability of the treatment being optimal among the current set of treatments in the domain [2], subject to these additional rules:

- The randomisation proportion for a given treatment cannot fall below 1 / (N + 2), where N is the number of treatments in the domain.
- The randomisation proportion for a given treatment cannot exceed (N + 1) / (N + 2)
- A new treatment added to the domain will be randomised at 1 / N for a burn-in period and all other treatments will be allocated using RAR in the remaining (N – 1) / N.

These randomisation proportions are set to maintain blinding to effect as clinicians administering the interventions cannot be blinded to the intervention.

Sample size

This study uses a perpetual design meaning that the sample size is not pre-defined. Domains will be stopped based on the stopping rules outlined in the stopping guidance section. Simulation methods will vary the number of individuals enrolled in the study to ensure good error control within each domain and to identify the number of patients that are likely to be required to draw sufficiently certain conclusions. Pilot/feasibility phases will have a fixed sample size, specified in the domain specific protocol.

Framework

Inferences in this trial will be based on a Bayesian statistical model, which estimates the posterior distribution of the parameter of interest, based on the accumulating data in the trial and pre-specified prior distributions. All trial conclusions will be based on the Bayesian posterior probability that a given treatment results in the best patient outcomes, measured by the domain-specific primary endpoint, compared to the current set of treatments randomized in that domain for individuals in that state. Study conclusions will be made separately for each state within each domain. However, where appropriate, we will use hierarchical modelling to dynamically integrate information across states to improve our ability to estimate the state-specific and domain-specific treatment effects. This borrowing can result in an increase in statistical power, while allowing for state specific treatment effects, and attenuate extreme results in smaller subgroups.

This framework differs from standard (frequentist) analyses where conclusions are based on assessing the likelihood of the observed data under a null hypothesis. This standard framework typically requires the specification of a control treatment that will be considered optimal, unless the trial data give evidence to the contrary. As our Bayesian framework is based on considering the probability that each treatment is superior compared to all other competing strategies in the domain, we do not designate a control treatment in the analysis and consider all treatments as equal until the response data are available. Based on the response data, a continuously updated ranking of the treatment regimens can be computed based on their effectiveness compared to all competing treatment regimens to which participants are currently being randomised.

Statistical interim analyses and stopping guidance

In our perpetual trial, all statistical analyses will be considered interim analyses. We will aim to undertake analyses every three months using data from all individuals who have completed data collection for the domain-specific primary outcome. Note that interim analyses may be skipped if fewer than 50 participants have been enrolled in the platform since the last analysis. Interim analyses may also be delayed for operational reasons relating to data availability. At each interim analysis, we will analyse the data for all the domain/state combinations that are currently randomizing patients and are not in their pilot/feasibility phase. The goal of these analyses is to determine whether statistical conditions (outlined below) have been met to indicate that the current treatments for the domain/state combination should be changed. Formally, the statistical rules outlined below offer guidance; final decisions regarding stopping and publication will be made by the Platform Steering Committee (PSC) and the Domain Protocol Committee (DPC) as applicable, under advisement of the DSMB. Each domain/state combination will have a pre-specified

minimum recruitment level before these statistical triggers are computed, which we can consider as a "burn-in" period for the study.

Statistical Triggers

The posterior distribution of the treatment effects will be estimated within each domain, separately for each state, using hierarchical models if appropriate (as defined by the domain specific protocol and SAP). For each state, we will compute the posterior probability that each treatment in the domain is superior, i.e., has the highest treatment effect among all treatments to which participants are currently being randomised. Treatments will be stopped in a specific domain and state combination, based on any of the following rules:

1) Stopped for superiority: randomisation will be stopped if one treatment has a high posterior probability of superiority, with the threshold specified in the domain specific protocol. All other treatments will be discontinued, and all study participants will receive this treatment until another treatment is considered for inclusion in the domain.
2) Stopped for inferiority: a specific treatment will be stopped for inferiority if the probability of superiority is small, as specified in the domain specific protocol. Note that a low posterior probability of superiority indicates a high posterior probability for harm.
3) Stopped for equivalence: if two treatments have a high posterior probability of being within the minimally important clinical difference (which we aim to harmonize across the domains with the same outcomes), then one of these treatments will be considered to stop for equivalence. If an equivalence trigger is met, the steering committee will determine which treatment is most suitable for stopping based on other data, e.g., safety.
4) Stopped for futility: futility can be declared in two manners: 1) Futility will be declared with respect to a given intervention, typically the standard or usual care and will be based on the probability that the treatment effect with respect to the standard intervention is less than a given threshold (e.g., a threshold larger than 1). Thus, this declaration of futility combines equivalence and inferiority. 2) The posterior probability of superiority is below a given threshold, where this threshold is larger than the inferiority threshold.

The specific stopping rules for each domain and intervention will be specified in the domain specific protocol and SAP. All stopping decisions will be made separately for each intervention within each domain for each state. Thus, a domain will continue enrolling eligible patients until the statistical trigger has been met separately for all states. However, patients will only be randomised in a domain in states for which the statistical trigger has not been met. Note that once a statistical trigger has been met and randomisation has ceased to a specific treatment, it will no longer be included in the interim analyses. Thus, treatments cannot be "reinstated" in the trial once the statistical trigger has been met and confirmed by the DSMB.

Timing of final analysis

Once a statistical trigger has been met for a domain and state combination, and the stopping decision confirmed by the DSMB and (PSC) and/or (DPC), we will proceed with the final analysis. Note that the PSC and/or DPC can decide to defer the final analysis in some settings, i.e., an inferiority trigger is met in a 3-arm domain, where the final analysis would be deferred until superiority is reached for one of the remaining two treatments. The final analysis will take place when all individuals randomized in this domain/state combination before the statistical trigger was confirmed by the DSMB and the PSC and/or DPC have completed their follow-up for all primary and secondary outcomes. As hierarchical modelling may be used, individuals who are not

randomized to the specific domain/state combination may be included in the final analysis. Thus, the data used for the final analysis in a specific domain/state combination will be all participants who have completed follow-up when the final participant randomised to the focal specific domain/state completes their follow-up for the primary outcome. Follow-up is considered complete for the purposes of this final analysis at the time the primary endpoint, as specified in the domain-specific protocol, should have become available based on the date of randomization, even if the value for the outcome is missing. This final analysis dataset will be locked and considered final for the given domain/state combination, even if the results could change based on the dynamic borrowing and updated outcomes in other states.

Timing of outcome assessments

Outcomes are collected in the iCORE, INDEX or I2 registries or in platform CRFs for sites not participating in a registry. Specific additional outcomes and their timing will be addressed in the domain specific protocols/SAPs.

Section 3: Statistical Principles

Statistical Reporting

All inferential statistics will be undertaken in a Bayesian framework [3]. Thus, when inferential statistics are used, we will report the posterior mean, median and 95% highest density credible interval for the parameter of interest. The primary analysis and pre-specified secondary analyses will also report the posterior probability of superiority (or its complement, inferiority). We will not report p-values and will not adjust for multiplicity as analyses are undertaken in a Bayesian framework [4]. Comprehensive simulations will be undertaken when a domain is added to ensure the proposed trial has good operating characteristics within each state.

Adherence and protocol deviations

Adherence and protocol deviations will be defined separately for each domain in the domain specific protocol and SAP. Adherence and protocol deviations will be presented using descriptive statistics by arm in each domain and state combination.

Analysis populations

Intention to Treat Population: All individuals who are randomised and meet the inclusion criteria and consent for the domain and state in which we are conducting the analysis. Individuals will be included in the intention to treat analysis irrespective of whether they received the treatment to which they were randomised. We may define some *delayed entry* domains, where individuals are only eligible for this domain after receiving other treatments, i.e., domains governing treatment in recovery. The intention to treat population for these domains will, similarly, include all individuals who meet the inclusion criteria and are consented for the domain. Thus, in delayed entry domains, the inclusion criteria will explicitly indicate the treatment and disease trajectory that must be followed to be eligible for delayed entry.

Per Protocol Population: The per protocol population will be a subset of the intention to treat population for a given state/domain combination. As the interventions are defined within each of the domain specific protocol, the per protocol population will be defined in the domain-specific protocol and SAP. Some domains may not include a per protocol population or analysis due to the complexity of the protocol and/or the interventions.

Safety: As the interventions are defined within each of the domain specific protocol, the safety population will be defined in the domain specific protocol and SAP.

Individuals enrolled in pilot/feasibility phases will be included in these analysis populations, provided the definition of the study outcomes does not change following the pilot/feasibility phase.

Section 4: Trial Population

Screening data

We will report number of patients screened who met inclusion criteria for the platform trial and the number excluded as they met one or more exclusion criteria. For each domain/state combination, we will then report the number of patients eligible for the state/domain combination and the number of patients randomized. Reasons why patients were eligible not-randomized will also be recorded. Finally, we will report the number of patients who received their randomized allocation.

Eligibility

Platform Inclusion Criteria

1. Acute hypoxemic respiratory failure meeting all of the following criteria;
   a. New or worsening respiratory symptoms developing within 2 weeks prior to the onset of need for oxygen or respiratory support
   b. Receiving any of the following types of oxygen or respiratory support for at least 4 hours prior to the time of randomization; supplemental oxygen at 10 L/min or higher, high flow nasal oxygen (at any flow rate), invasive ventilator support, extra-corporeal life support (ECLS), or non-invasive ventilator support
   c. Minimum FiO2 $\geqslant$ 0.40 (for venturi mask, high flow nasal cannula, or invasive or non-invasive ventilation) or oxygen flow rate $\geqslant$10 L/min on face mask for at least 4 hours at the time of evaluation for eligibility unless already on extra-corporeal life support
2. Age $\geqslant$ 18 years
3. Hypoxemia not primarily attributable to acute heart failure, fluid overload, or pulmonary embolism (PE)

Platform Exclusion Criteria

1. Extubation is planned or anticipated on the day of screening
2. ICU discharge is planned or anticipated on the day of screening
3. The patient is moribund and deemed unlikely to survive past 24 hours (as determined by the clinical team)
4. The patient is being transitioned to a fully palliative philosophy of care

Each state/domain combination may have additional inclusion and exclusion criteria, which will be specific in domain protocols according to domain-specific requirements.

Platform Defined States

The platform structure defines multiple mutually exclusive patient states assessed at the time of randomization in a given domain. The pre-specified states for the platform include:

1. non-intubated patients;
2. intubated patients, not on extracorporeal life support, with low normalized respiratory system elastance (<2.5 cm H2O/ml/kg predicted body weight);
3. intubated patients, not on extracorporeal life support, with high normalized respiratory system elastance (≥2.5 cm H2O/ml/kg predicted body weight);
4. patients on extracorporeal life support.

Recruitment

We will use a CONSORT 2010 flow diagram to present the recruitment numbers within each domain/state combination. In addition to the screening data explained above, the CONSORT diagram will detail the number of patients who were lost to follow-up, discontinued the intervention were included/excluded from the primary analysis.

Withdrawal/follow-up

Patients can withdraw from the platform at any time and for any reason, upon request. The reason for voluntary withdrawal from the study will not be collected. Participants who choose to withdraw from the study can choose to either discontinue the study intervention but have their outcomes recorded or withdraw from all further data collection. Individuals who discontinue the study intervention but have their outcomes recorded will be included in the intention to treat analysis. All reasons for withdrawal will be tabulated and we will record whether the patient consented to further follow-up. The numbers (with reasons) of losses to follow-up over the course of the trial will be summarised by treatment arm for each domain and within each state.

Baseline patient characteristics

We will report the following baseline characteristics for all domains:

1) Age
2) Sex
3) Diagnostic reason for ICU admission
4) Respiratory support at randomization
5) APACHE
6) SOFA
7) P/F ratio
8) Ventilatory Ratio

We may also report domain specific baseline characteristic that will be specified in the domain specific protocol.

Our presentation of these baseline characteristics will depend on the type of data. Categorical variables will be compared between the treatment groups using frequency counts and percentages. Continuous variables will be presented using the mean, median, standard deviation and interquartile range separately for each treatment group. Baseline characteristics will be reported separately for each domain/state combination.

Section 5: Analysis

Outcome definitions

Because this platform is intended to host trials focused on various aspects of acute hypoxemic respiratory failure at different phases of investigation, we cannot specify a single primary

endpoint for the entire platform. Rather, the primary endpoint will be defined by each domain protocol. Similarly, secondary endpoints of interest will be defined in domain protocols.
In general, irrespective of domain phase and interventions, all domains of PRACTICAL will be expected to ascertain the following key endpoints:

Clinical endpoints

- Survival status (dead or alive) at hospital discharge, defined as the date/time the patient is discharged from the study hospital to another location outside the study hospital
- Survival status (dead or alive) at ICU discharge, defined as the date/time the patient is discharged from the study ICU to another location within or outside the study hospital.
- ICU length-of-stay measured from the day of randomisation until the day of ICU discharge (defined above).
- Survival status (dead or alive) at disconnection from mechanical ventilation, defined as the final date of extubation or the first day of continuous tracheostomy mask for at least 24 hours, provided ventilatory support is not resumed during the index ICU admission.
- Duration of mechanical ventilation measured from the day of randomisation until the day of disconnection from mechanical ventilation (defined above).
- Ventilator-free days to day 28, which is an ordinal scale ranging from –1 to 29, defined as:
    - Patients who die in hospital on or before day 90 will be assigned a score of –1
    - Patients who are not assigned –1 but are not liberated from invasive ventilation or extracorporeal life support by day 28 will receive a score of 0
    - Patient who are not assigned –1 and who are on ventilatory support at baseline will be assigned a score computed from the number of days alive and free of respiratory support from the day of randomization to day 28 after randomization
    - For this calculation, a day is counted as a calendar day, from midnight to midnight
        - Day of randomization is counted as day 0
        - Day 28 is 28 calendar days after day 0
        - A day “alive and free of ventilation” is a day on which the patient does NOT require non-invasive mechanical ventilation OR invasive mechanical ventilation, OR extracorporeal life support for ANY duration at ANY time during that calendar day
        - The status on the day of randomization is counted toward this computation
        - A patient who is extubated on the day of randomization would receive a score of 28
        - A patient who is extubated on the day after randomization would receive a score of 27
    - Patients who are not assigned –1 and who are not on ventilatory support at randomization and who never require ventilatory support (defined as non-invasive ventilation OR invasive ventilation OR extracorporeal life support) up to day 28 will be assigned a score of 29
    - For patients who are not assigned –1 and who are not on ventilatory support at randomization and who eventually require ventilatory support, the days alive and free of ventilation prior to the initiation of ventilation will not count toward the computation of VFDs

        - i.e. A patient who is not on ventilatory support at baseline but is intubated on day 3 and liberated on day 12 and survives to hospital discharge will have a ventilator-free day score of 16.
    - For patients in whom non-invasive mechanical ventilation AND invasive mechanical ventilation AND extracorporeal life support is discontinued but then subsequently resumed at any time on or before day 28, the days alive and free of invasive ventilation and ECLS in the intervening period will NOT count as ventilator free days (only the days free of ventilation all the way to the end of day 28 will count; i.e., if a patient is liberated from ventilation on day 2 and restarts on day 5 before finally ending on day 10, their VFD outcome will be 18 and not 21).

Safety endpoints: (The clinical endpoints defined above also function as safety endpoints)

- For non-intubated patients, change from non-intubated to intubated or ECLS
- For intubated patients, change from intubated to ECLS
- Serious adverse events (SAE) will be defined in each domain protocol.

Individuals can transition between disease states throughout the trial but will be included in the analysis for the state that they were in when they were randomised. All outcomes will be measured from the time of randomization into the domain. Additional pilot/feasibility outcomes for domain in this phase will be defined in the domain specific protocols, harmonized where possible. All additional outcomes will be defined in the domain specific protocols and SAPs.

Analysis methods

The specific outcome model will change depending on the outcome type. Thus, this section describes the general analysis approach. Statistical modelling will adjust for variation across sites, trial time period (in 2-week intervals), the stratification variables used in randomisation (if applicable), sex, age and potential domain specific variables. Adjustments will be made using a linear predictor framework, where the link function and error distribution will change depending on outcome type. The state-specific treatment effect for the randomised treatments within a domain will be estimated on the linear predictor scale and will be additive. When pre-specified in the domain specific SAPs, baseline effects and interaction effects between treatments in different domains will be estimated in the same model. Interaction effects will only be considered when treatments for the different domains are offered concurrently. The primary analysis will not consider treatment-covariate interactions, although pre-specified secondary analyses could consider these interactions and will be specified in the domain specific analysis plans. Hierarchical modelling will be used, where pre-specified, to borrow information about the treatment effect across states.

The specific treatment effect reported will change depending on the model used. We will endeavour to present treatment effects on an interpretable scale, i.e., hazard ratios rather than log hazard ratios. The mean treatment effect and 95% high density posterior credible interval will be calculated directly from the posterior distribution of the parameter from the regression. Thus, due to the flexibility of Bayesian modelling, we hope to report functions of the parameters estimated using the linear predictor framework and ensure interpretability of our findings.

Key modelling assumptions such as linearity, proportional hazards and the assumed error distributions will be checked graphically before model fitting. Where appropriate, we will also evaluate residual plots. These checks will depend on the outcome model and will be specified in the domain specific SAP.

In general, we will use minimally informative priors for all analyses, using half-Cauchy priors for standard deviations [5] and centered t-distributions for coefficients [6]. We will undertake prior predictive checks to ensure that these minimally informative priors are jointly minimally informative [7]. In domains and states with low enrollment numbers, we will consider using informative priors obtained from previous studies, downweighed to account for differences in historical data, or elicited from experts who are not involved in the PRACTICAL study. Where informative priors are used, we will undertake a sensitivity analysis by performing the analysis with minimally informative priors. The specific priors will be defined in the domain specific SAPs.

Subgroup analyses will be pre-specified in the domain specific protocols. Subgroup analyses will be implemented using the same model structure as used in the primary analysis with a main effect for the subgroup of interest and an interaction effect with the treatment. Priors will be specified as above and the results will be reported using the mean effect, 95% credible interval and posterior probability of a positive interaction effect between treatment and the subgroup.

Pilot/Feasibility Phases
Pilot/feasibility outcomes will be summarised using descriptive statistics with no formal inferential statistics used. The descriptive statistics used will depend on the outcome type. In general, Categorical variables will be reported using frequency counts and percentages and continuous variables will be reported using the mean, median, standard deviation and interquartile range. Pilot/feasibility outcomes may be reported separately by treatment group, where applicable.

Missing data

Descriptive statistics on the proportion of missing data will be provided, identifying the proportion of missing data in the outcomes and the covariates.

Covariates
If the level of missingness for the covariates is below 5%, we will undertake a complete case analysis using the observed data. If the level of missingness exceeds 5%, we will evaluate whether a complete case analysis is suitable by investigating the baseline characteristics between individuals with and without the missing outcome. Using this information, expert opinion will be leveraged to determine whether a complete case analysis could be appropriate in this setting. If a complete case analysis is not suitable, we will undertake a joint Bayesian analysis of the missingness mechanism and outcome model. This joint analysis will be the primary analysis. We will also undertake a secondary complete case analysis to explore any differences.

Outcomes
In general, we will not jointly model or impute missing outcomes. Thus, if the level of missingness for a specific outcome exceeds 5%, we will investigate the baseline prognostic characteristics and treatment assignment of individuals with and without the missing outcome. If substantial differences are observed, we will report descriptive statistics for this outcome and report the

reasons that individuals were lost to follow up. If we are analysing a composite outcome, we will consider joint modelling of the missingness mechanism and specific elements that contribute to the outcome. We would jointly model the missingness for the outcomes that contribute to the composite outcome, if the missingness for that specific element exceeds 5%, and the outcome.

Additional analyses

Additional analyses will be defined in the domain specific protocols and may include joint modelling of different outcomes. Inferential statistics will follow the general principles outlined in this SAP.

Harms

Safety outcomes will be specified in the domain specific protocols. Unless specifically specified in the domain specific SAP, inferential analyses will not be used for safety data. Descriptive statistics detailing the number and proportion of adverse events observed in each category will be provided. The safety population and the safety data definitions will be defined in the domain specific protocols.

Statistical software

Analysis will be undertaken in specialised Bayesian software. Where possible, we will use the Integrated Nested Laplace Approximation (INLA) method to fit models through the R-INLA package due to its relative efficiency compared to sampling-based algorithms [8]. Where this algorithm is not suitable, e.g., for multinomial outcomes, we will use Markov Chain Monte Carlo (MCMC) methods through software packages such as *jags* or *stan* [9, 10]. Primarily, these will be accessed through R [11]. Additional analyses may be completed in SAS or Excel. Trial simulations to assess trial error rates will either be undertaken in R, interfacing with software to performing Bayesian updating, or in specialised software such as FACTS (Fixed and Adaptive Clinical Trial Simulator).

Bibliography

[1] D. Angus, S. Berry, R. Lewis, F. A. Y. Al-Beidh, W. van Bentum-Puijk, Z. Bhimani, M. Bonten, K. Broglio, F. Brunkhorst and A. Cheng, "The REMAP-CAP (randomized embedded multifactorial adaptive platform for community-acquired pneumonia) study. rationale and design.," *Annals of the American Thoracic Society,* vol. 17, no. 7, pp. 879-891, 2020.

[2] J. Wathen and P. Thall, "A simulation study of outcome adaptive randomization in multi-arm clinical trials," *Clinical Trials,* vol. 14, no. 5, pp. 432-440, 2017.

[3] A. Gelman, B. Carlin, H. Stern, D. Dunson, A. Vehtari and D. Rubin, Bayesian Data Analysis [3rd Edition], Boca Raton, FL: CRC Press, 2013.

[4] D. Berry and Y. Hochberg, "Bayesian perspectives on multiple comparisons.," *Journal of Statistical Planning and Inference,* vol. 82, no. 1-2, pp. 215-227, 1999.

[5] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)," *Bayesian analysis,* vol. 1, no. 3, pp. 515-534, 2006.

[6] A. Gelman, A. Jakulin, M. Pittau and Y. Su, "A weakly informative default prior distribution for logistic and other regression models," *The annals of applied statistics,* vol. 2, no. 4, pp. 1360-1383, 2008.

[7] A. Gelman and C. Shalizi, "Philosophy and the practice of Bayesian statistics," *British Journal of Mathematical and Statistical Psychology,* vol. 66, no. 1, pp. 8-38, 2013.

[8] H. Rue, A. Riebler, S. Sørbye, J. Illian, D. Simpson and F. Lindgren, "Bayesian computing with INLA: a review," *Annual Review of Statistics and Its Application,* vol. 4, no. 395-421, pp. 395-421, 2017.

[9] A. Gelman, D. Lee and J. Guo, "Stan: A probabilistic programming language for Bayesian inference and optimization.," *Journal of Educational and Behavioral Statistics,* vol. 40, no. 5, pp. 530-543, 2015.

[10] M. Plummer, " JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," *In Proceedings of the 3rd international workshop on distributed statistical computing,* vol. 124, no. 125.10, pp. 1 - 10, 2003.

[11] R Core Team , R: A language and environment for statistical computing, Vienna. Austria: R Foundation for Statistical Computing, 2020.

Appendix: Inclusions for Domain-Specific Protocols/Analysis Plans

Randomization
Stratification Variables
Response Adaptive Randomisation
- Will RAR be used?
- Burn-in period

Sample Size
Feasibility study sample size

Statistical interim analyses and stopping guidance
Minimum recruitment level before statistical triggers are computed.
Probability threshold to stop for superiority
Probability threshold to stop for inferiority
Probability threshold to stop for equivalence
Minimally important clinical difference for equivalence
Futility analysis
- Will a futility analysis be used?
- Which treatments will be subject to a futility analysis?
- What probability threshold will be used?

Timing of outcome assessments
Timing of collection for all outcomes.

Adherence and protocol deviations
Definition of adherence
Definition of a protocol deviation

Analysis populations
Definition of per-protocol population
Definition of safety population

Eligibility
Domain specific inclusion/exclusion criteria

Baseline patient characteristics
Domain specific baseline characteristics to report

Outcome Definitions
Primary outcome
Secondary outcomes
Feasibility outcomes

Analysis Methods
Domain specific adjustment variables

Error distribution and link function for each outcome
Potential treatment interactions with other domains
Potential treatment covariate interactions for secondary analyses
Hierarchical model structure across states or domains
Model checking procedures
Informative priors

- Will they be used?
- How will they be defined

Subgroup analyses
Additional analyses
Safety data analyses