Simulation report: Main results

Tailored Interventions Guided by Endotypic Responses in Sepsis (TIGERS)

Date of report: 21 November 2025

Version 1.2

Simulation report produced by:
Dr Anais Charles-Nelson
Australia and New Zealand Intensive Care Research Centre (ANZIC-RC)
School of Public Health and Preventive Medicine
Monash University
Melbourne

Dr Elizabeth Ryan
Biostatistics Unit and Monash University Clinical Trials Centre
School of Public Health and Preventive Medicine
Monash University
Melbourne

1 Version history

Version	Date	Description	Authors	Software
V1.0	09/07/2025	Simulation results for the grant application	Dr Anais Charles-Nelson Dr Elizabeth Ryan	R version 4.4.3 (2025-02-28 ucrt) INLA_24.12.11 built 2024-12-11 19:58:26 UTC
V1.1	30/10/2025	Incorporated overall summary tables for relevant scenarios (in the executive summary); incorporated updated simulation results (10,000 simulations for certain "power" scenarios); minor modifications to the text elsewhere.	Dr Anais Charles-Nelson Dr Elizabeth Ryan	R version 4.4.3 (2025-02-28 ucrt) INLA_24.12.11 built 2024-12-11 19:58:26 UTC
V1.2	21/11/2025	Added results of additional simulations (for SRS1) to the executive summary, where we varied the recruitment assumptions	Dr Anais Charles-Nelson Dr Elizabeth Ryan	R version 4.4.3 (2025-02-28 ucrt) INLA_24.12.11 built 2024-12-11 19:58:26 UTC

Table of contents

List of figures

List of tables

2 Scope of the report

This simulation report describes the simulation exercise that was performed for the TIGERS platform trial and presents the results. The purpose of this report is to outline the design and adaptive rules applied, describe the simulation methods used, and summarise the operating characteristics derived from the simulation exercise for the TIGERS platform trial.

More specifically, this simulation report aims to address the following objectives for the TIGERS grant application:

- To identify appropriate values for the statistical triggers (futility, efficacy stopping).
- To determine the platform's operating characteristics (type I error, power, sample size and time taken to evaluate treatments), using the identified statistical triggers, under various assumed scenarios regarding recruitment / treatment effects.

3 Executive summary

For this initial set of interventions that we will investigate in the TIGERS platform trial for the SRS 1 and SRS 2 endotypes/subtypes, we will recruit a maximum of 450 participants per intervention arm in the SRS 1 endotype, and 400 per intervention arm in the SRS 2 endotype. (Note that the numbers recruited to the control arm may exceed these caps in each endotype). Provided there is no early stopping in any of the intervention arms, the maximum sample size that we would recruit in SRS 1 is 1185 participants, and the maximum that we would recruit in the SRS 2 endotype is 1287 participants, giving a total of 2472 participants (assuming non-binding futility), for an assumed recruitment rate of 1.5 patients/site/month and duration (42 months recruitment). (Note that these numbers would increase if we assumed a higher recruitment rate, or longer duration (see results at end of section 3)).

Under our target power scenario (POR = 1.37 for both intervention arms in SRS1, and POR = 1.42 for both intervention arms in SRS2, representing a 5% and 4.5% absolute decrease in mortality, respectively, and a change of 3.5 and 4.5 days (respectively) in the median OSFD), we would expect an average sample size of 735 for SRS 1 (across all three arms) and an average sample size of 771 in SRS 2 (across all three arms), assuming non-binding futility.

If we assume binding futility we would have a maximum of 1147 participants in SRS1 and 1281 participants in SRS2 (under the scenarios that we explored). Under our target power scenario, we would expect an average sample size of 672 for SRS 1 (across all three arms) and an average sample size of 733 in SRS 2 (across all three arms), assuming binding futility.

In the table below, we present the final design settings that were used for the TIGERS platform trial primary simulations. These were obtained after several rounds of simulations which investigated different interim analysis schedules, recruitment rates and durations, sample size caps, primary outcome distributions, treatment effect scenarios, and efficacy and futility rules. The efficacy rules were chosen to ensure (pairwise) 1-sided type I error of approximately 10% and (pairwise) power ~80% (assuming non-binding futility), and were chosen in consultation with the clinical investigators to ensure that clinically-relevant thresholds were used. The futility rules incorporated the same clinically-important treatment effect as the

efficacy stopping rules (POR_{fut} = POR_{eff}), and the threshold (p_{fut}) was chosen to ensure that appropriate amounts of arm dropping for futility occurred when a small or non-beneficial treatment effect was assumed.

Parameter	Main
Recruitment	• 42 months duration • 1.5 patients per site per month • 60 sites total (takes 21 months to get all sites up and running) • Approx. max. 2472 patients in total (across all treatment and groups under the null scenario with non-binding futility)
Sample size cap	450 in SRS 1 and 400 in SRS2 for each intervention
Treatment effect size (POR for Anti-IL6, POR for anti-TNF alpha)	• (1,1) (type I error) • (1.37,1.37) (power for SRS1) • (1.42,1.42) (power for SRS2)
Split between phenotype	40% (SRS 1) vs 60% (SRS 2)
Number of patients in each arm to trigger the first interim analysis	50
Statistical triggers at first interim analysis	Futility only (non-binding)
Time of the subsequent interim analyses	6-monthly
Efficacy stopping rule	P(log(POR)>log(1.1))>0.83 (applied at interim analysis 2 onward
Futility stopping rule	P(log(POR)<log(1.1))>0.8 (applied at each interim analysis; non-binding)

POR = Proportional Odds Ratio

The operating characteristics for each endotype subtrial are displayed in the table below (for certain effect size scenarios). The one-sided type I error was approximately 10% for each comparison of the intervention arms to the control (pairwise type I error) for each SRS endotype. Assuming a POR = 1.37 for an intervention arm in SRS1, the power to declare efficacy of each intervention (compared to control) was ≥75% (pairwise power), and the probability of declaring efficacy for at least one of these interventions is >90% (when both interventions are assumed to have POR=1.37; "any pair" power). For SRS2, assuming a POR = 1.42 for an intervention arm the (pairwise) power was > 83% and the probability of declaring efficacy for at least one of these interventions ("any pair" power) is >95% (when both interventions are assumed to have POR=1.42). There was a low probability of declaring superiority when a harmful

treatment effect (POR=0.9) was assumed (3.4% in SRS1 and 1.6% in SRS 2). These results were obtained under the assumption of non-binding futility.

Under the assumption of binding futility, the type I error was slightly decreased and the power was ≥72% for SRS1 and ≥ 80% for SRS2 under the target POR. Truly effective arms were dropped <12% of the time (assuming binding futility rules).

Operating characteristics for each endotype sub-trial for certain scenarios (assuming 1.5 patients/site/month over 42 months)

Scenario	Average/Expected sample size (range)	Expected duration in months (range)	Prob declare efficacy treatment 1	Prob declare efficacy treatment 2	Prob efficacy declared for at least one intervention*	Prob early stop treatment 1 for efficacy	Prob early stop treatment 2 for efficacy	Prob early stop treatment 1 for futility**	Prob early stop treatment 2 for futility**	Overall prob of inconclusive result in subtype***
SRS 1										
Non-binding futility										
Null: POR1=1, POR2 = 1	1031 (373 – 185)	44 (25 – 44)	0.1038	0.1099	0.1861	0.1434	0.1492	NA	NA	0.2476
Both work: POR1 = 1.37, POR2 = 1.37	735 (365-1165)	36 (25 – 44)	0.7688	0.7545	0.9048	0.7855	0.7862	NA	NA	0.0879
One works: POR1 = 1.37, POR2 = 1	982 (373 – 182)	42 (25 – 44)	0.7492	0.1080	0.7592	0.7740	0.1454	NA	NA	0.0857
Binding futility										
Null: POR1 = 1, POR2 = 1	741 (190-1147)	36 (20 – 44)	0.0956	0.103	0.1700	0.1467	0.1467	0.5849	0.5920	0.1710
Both work: POR1 = 1.37, POR2 = 1.37	672 (190-1125)	34 (20-44)	0.7182	0.7434	0.8919	0.7457	0.7770	0.1150	0.1029	0.0535
One works: POR1 = 1.37, POR2 = 1	744 (190-1147)	36 (20-44)	0.7252	0.0974	0.7353	0.7623	0.1460	0.1108	0.5861	0.0700
SRS 2										
Non-binding futility										
Null: POR1 = 1, POR2 = 1	1198 (424 – 1287)	37 (23 – 43)	0.0916	0.0920	0.1642	0.1355	0.1363	NA	NA	0.2361

Both work: POR1 =1.42, POR2 =1.42	771 (412-1281)	29 (22-43)	0.8390	0.8436	0.9530	0.8273	0.8283	NA	NA	0.0454
One works: POR1 = 1.42, POR2 = 1	1016 (424-1282)	34 (23-43)	0.8370	0.0867	0.8422	0.8683	0.1241	NA	NA	0.0492
Binding futility										
Null: POR1 = 1, POR2 = 1	796 (197-1279)	30 (17 – 40)	0.0782	0.0844	0.1379	0.1146	0.1238	0.5762	0.5865	0.1780
Both work: POR1 =1.42, POR2 =1.42	733 (201 – 1281)	29 (17-40)	0.7971	0.7848	0.9279	0.7942	0.7685	0.0730	0.0802	0.0412
One works: POR1 = 1.42, POR2 = 1	780 (201 – 1271)	30 (17 – 39)	0.8128	0.0817	0.8211	0.8115	0.1184	0.0711	0.5623	0.0508

POR1 = proportional odds ratio for investigational treatment 1 (e.g., tocilizumab);

POR2 proportional odds ratio for investigational treatment 2 (e.g., infliximab)

*Proportion of simulated trials that the efficacy rule was met for at least one of investigational treatment 1 or treatment 2

**Binding futility designs only; NA for designs that implemented non-binding futility rules

***An inconclusive result occurs if neither the efficacy or futility rules have been met (in either investigational arm)

A number of different effect size scenarios were explored, where 1000 trials were simulated for each parameter setting. For certain scenarios, such as the "null scenario" (for estimating type I error) and the "both work" or "one works" (for estimating power), 10 000 trials were simulated so that more precise estimates of the operating characteristics could be obtained for these scenarios.

Sensitivity analyses were conducted that explored different recruitment rates, primary outcome distributions and SRS1:SRS2 ratios; these are presented in a separate document (as supplementary material; briefly, it appears as though the type I error remains controlled at approx. 10% (1-sided) and similar power (to that described above) is obtained for our target POR when we alter our assumptions regarding the above-mentioned parameters).

The primary simulation results assumed non-binding futility (i.e., the futility stopping rules were "turned off" / "not obeyed"), but we also performed simulations assuming binding futility (futility rules were always obeyed). In some instances, the type I error (1-sided) is slightly above the target of 10%

We also performed some additional simulations to determine if power could be improved for the SRS1 subtype if the sample size were increased slightly. First, we performed some simulations which explored the type I error and power scenarios if we were to recruit at a slightly higher recruitment rate of 1.6 patients/site/month (across 42 months and 60 sites). Under these assumptions, we obtained a slight increase in power:

Operating characteristics for SRS1 if increase recruitment to 1.6 patients/month in 60 sites (42 months)

Scenario	Expected sample size (range)	Expected duration in months (range)	Prob declare efficacy treatment 1	Prob declare efficacy treatment 2	Prob efficacy declared for at least one intervention*	Prob early stop treatment 1 for efficacy	Prob early stop treatment 2 for efficacy	Overall prob of inconclusive result in subtype**
Null: POR1 = 1, POR2 = 1	1090 (398 – 1236)	43 (25 – 44)	0.1022	0.1084	0.1779	0.1420	0.1490	0.2290
Both work: POR1 = 1.37, POR2 = 1.37	759 (385 – 1236)	35 (25 – 44)	0.7812	0.7853	0.9000	0.8063	0.7952	0.0859

POR1 = proportional odds ratio for investigational treatment 1 (e.g., tocilizumab);

POR2 proportional odds ratio for investigational treatment 2 (e.g., infliximab)

*Proportion of simulated trials that the efficacy rule was met for at least one of investigational treatment 1 or treatment 2

**An inconclusive result occurs if neither the efficacy or futility rules have been met (in either investigational arm)

We also explored extending recruitment to 66 sites and to 48 months (with 1.5/site/month) and obtained an increase in power.

Operating characteristics for SRS1 if increase recruitment to 66 sites across 48 months (1.5/site/month)

Scenario	Expected sample size (range)	Expected duration in months (range)	Prob declare efficacy treatment 1	Prob declare efficacy treatment 2	Prob efficacy declared for at least one intervention*	Prob early stop treatment 1 for efficacy	Prob early stop treatment 2 for efficacy	Overall prob of inconclusive result in subtype**
Null: POR1 = 1, POR2 = 1	1289 (377 - 1436)	48 (26 – 50)	0.1102	0.1020	0.1854	0.1433	0.1382	0.2000
Both work: POR1 = 1.37, POR2 = 1.37	795 (377 – 1425)	36 (25 – 50)	0.7969	0.7926	0.9319	0.8254	0.8193	0.0660

4 Background

Sepsis is a complex and heterogeneous syndrome caused by a dysregulated immune response to infection, leading to organ failure and death. It is the leading cause of intensive care unit (ICU) admissions and is responsible for one in five deaths globally. Despite decades of research, no pharmacological treatments beyond antibiotics have been proven to improve outcomes.

Traditional approaches to sepsis have relied on broad syndromic definitions, which overlook important biological differences between patients. This has contributed to the failure of many clinical trials and limited the translation of promising preclinical findings into effective therapies. A new approach is emerging that focuses on identifying sepsis endotypes—biologically distinct subgroups defined by differences in immune response—rather than treating all sepsis cases as the same.

Recent studies show that these endotypes respond differently to treatments, as demonstrated by secondary analyses of completed trials. This suggests that matching therapies to specific immune profiles may be the key to successful intervention.

This project will deliver an innovative, precision medicine-based, phase 2 adaptive platform trial to test therapies aimed at modifying the dysfunctional immune responses that drive sepsis-related mortality. This area represents a critical unmet need and has been identified as a top research priority by multiple national and international stakeholders.

NB: In this report we use the terms "subtype", "subgroup", "endotype" and "subphenotype" interchangeably to refer to the SRS 1 and SRS 2 groups.

5 Study objectives and endpoints

5.1 Study objectives

5.1.1 Overall aim

To accelerate the development of pharmacological therapies for sepsis by establishing an international phase 2 precision medicine adaptive platform trial.

5.1.2 Initial objectives

1. To test in patients with sepsis with SRS1/SRS2 endotypes the efficacy of tocilizumab or infliximab to improve 28-day organ support-free days.
2. To develop an infrastructure for testing additional endotypes and therapies in a platform trial.

5.1.3 Research question

In patients with sepsis with the SRS 1 or 2 endotype, do immune modulating drugs in addition to usual care, compared to usual care alone, reduce the risk of death or the duration of organ support?

5.2 Endpoints

5.2.1 Primary endpoint

The primary outcome is a 30-category ordinal variable combining days free of organ support up to day 28 for survivors (values 0 to 28) with death (value -1). Organ support can be either i) Respiratory support, defined as invasive mechanical ventilation (IMV) or non-invasive ventilation (NIV) including continuous positive airway pressure (CPAP) or high flow nasal oxygen (HFNO) with a prescribed FiO2 ≥ 0.4 and HFNO rate ≥ 30L/min or ii) Cardiovascular support, defined as the continuous infusion of any vasopressor or inotrope medication or iii) renal support (defined as renal replacement therapy). Organ support that is normally received by the patient outside of the hospital (e.g. CPAP or dialysis in the community) does not count towards this outcome (only new organ support for that ICU admission).

5.2.2 Secondary endpoints

- Ordinal scale comprising 90-day mortality combined with clinical state (in-hospital with organ support, in-hospital without organ support, discharged from hospital) over time,
- 28-day vasopressor-free days,
- 28-day respiratory support–free days,
- 28-day renal replacement therapy-free days,
- ICU length of stay,
- Hospital length of stay,
- All-cause mortality at 28 and 90 days,

- Serious adverse events,
- Health-related quality of life at 90 days (EQ-5D-5L).

6 Study methods

6.1 Design

6.1.1 General study design and plan

Participants will be recruited from multiple sites internationally. The TIGERS platform will use a Bayesian Adaptive Multi-Arm Trial design, which can be viewed as an extension of a Multi-Arm Multi-Stage (MAMS) design.

TIGERS is a randomised, open-label Bayesian adaptive multi-arm platform trial with pre-defined triggers for efficacy and futility stopping (as compared to usual care). Randomisation will be balanced with equal ratios between usual care and each intervention that participants are eligible for, i.e. a 1:1 ratio (usual care versus intervention (where eligible) by subphenotype). There is no fixed sample size but we have capped the sample size for the initial treatments and subphenotype. We will recruit a maximum of 450 per active intervention in the SRS 1 subphenotype, and 400 per active intervention in the SRS 2 subphenotype.

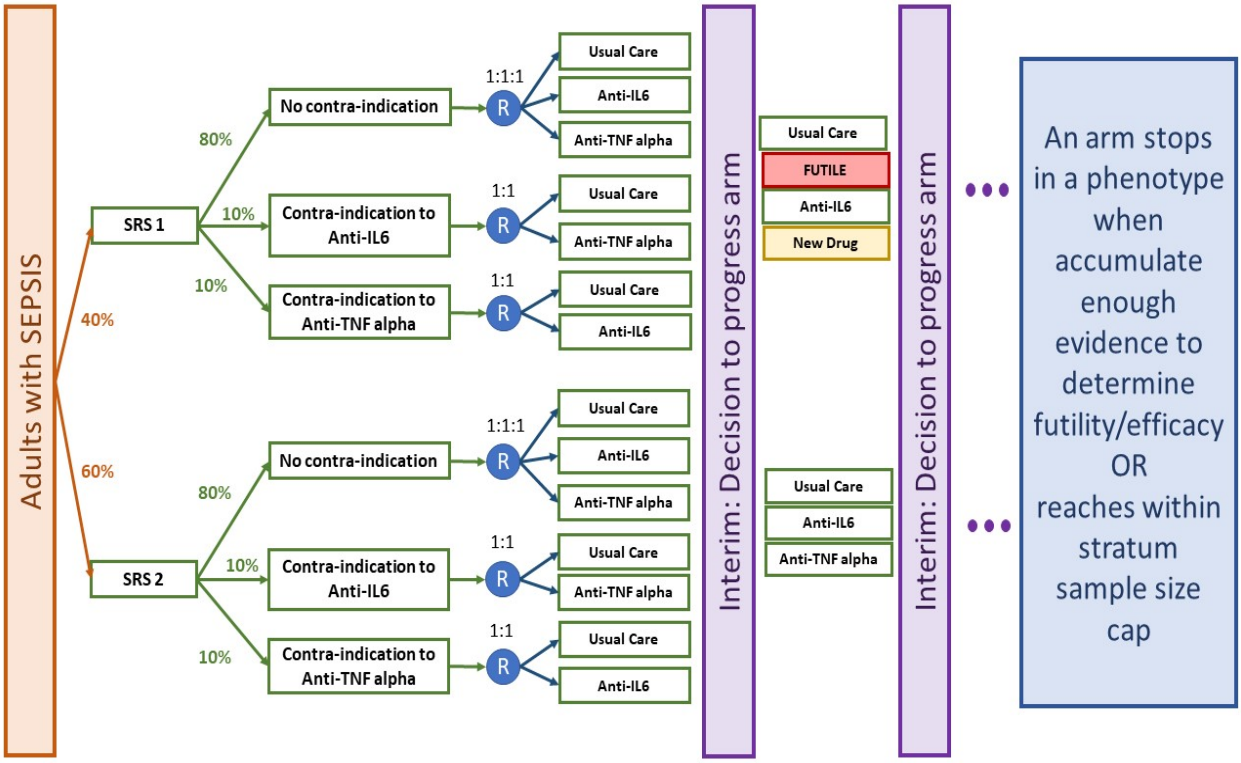

FIGURE 1: DESIGN

6.1.2 Randomisation and blinding

Treatment allocation will use a dynamic minimisation algorithm based on participant distribution within phenotype, country, and intervention eligibility. The allocation process will be managed through a secure online system, coordinated by the IMPERIAL Trials Unit. The algorithm will be updated as phenotype, countries, or interventions change. Clinical staff will be blinded to phenotyping results. However, due to challenges and costs in blinding for multi-arm trials, an open-label design was selected, as has been used in other successful adaptive platform trials in critical care (e.g., RECOVERY, REMAP-CAP), using objective outcomes (e.g., organ support-free days, all-cause mortality) that are unaffected by knowledge of treatment allocation. During the ongoing TIGERS platform, only unblinded trial statisticians will have access to emerging aggregate data and treatment allocation results until interventions are stopped. These statisticians will perform the adaptive analyses and generate statistical reports for regular DMC meetings. The unblinded statisticians conducting adaptive analyses will not participate in the platform's design decisions.

6.1.3 Sample size cap

We will recruit a maximum of 450 per intervention arm in the SRS 1 endotype, and 400 per intervention arm in the SRS 2 endotype. (Note that the numbers recruited to the control arm may exceed these caps in each endotype). Provided there is no early stopping in any of the intervention arms, the maximum sample size that we would recruit in SRS 1 is 1185 participants, and the maximum that we would recruit in the SRS 2 endotype is 1287 participants, giving a total of 2472 participants (assuming non-binding futility), for our assumed recruitment rates (1.5/site/month) and duration (42 months). Note that if we were able to recruit longer, or at a faster rate, these values would increase (See end of Section 3).

Under our target power scenario (POR = 1.37 for both intervention arms in SRS1, and POR = 1.42 for both intervention arms in SRS2), we would expect an average sample size of 735 for SRS 1 (across all three arms) and an average sample size of 771 in SRS 2 (across all three arms), assuming non-binding futility.

If we assume binding futility we would have a maximum of 1147 participants in SRS1 and 1281 participants in SRS2. Under our target power scenario, we would expect an average sample size of 672 for SRS 1 (across all three arms) and an average sample size of 733 in SRS 2 (across all three arms), assuming binding futility.

6.1.4 Decision rules and adaptations

6.1.4.1 Interim analysis timing

The first interim analysis will be performed after the first **50 patients** are due to have completed follow-up in each of the three arms (in an endotype). At the first interim analysis, **only early stopping for futility may occur.** The subsequent interim analyses will be conducted on a 6-monthly basis, where either early stopping for efficacy or futility may occur.

6.1.4.2 *Early stopping rules*

Arm dropping (of an investigational treatment) is allowed throughout the study if it meets one of the stopping rules:

- Futility: if $P(\log(POR)<\log(POR_{fut}))>p_{fut}$ where $POR_{fut}=1.1$ and $p_{fut}=0.8$
- Efficacy: if $P(\log(POR)>\log(POR_{eff}))>p_{eff}$ where $POR_{eff}=1.1$ and $p_{eff}=0.83$

These values were chosen to produce a (pairwise) 1-sided type I error of approximately 10%, and a pairwise power > 70% (several rounds of simulations were performed to obtain these values). The POR_{fut} and POR_{eff} were also chosen based on clinical investigator input.

An illustration of these stopping rules is given in Figure 2.

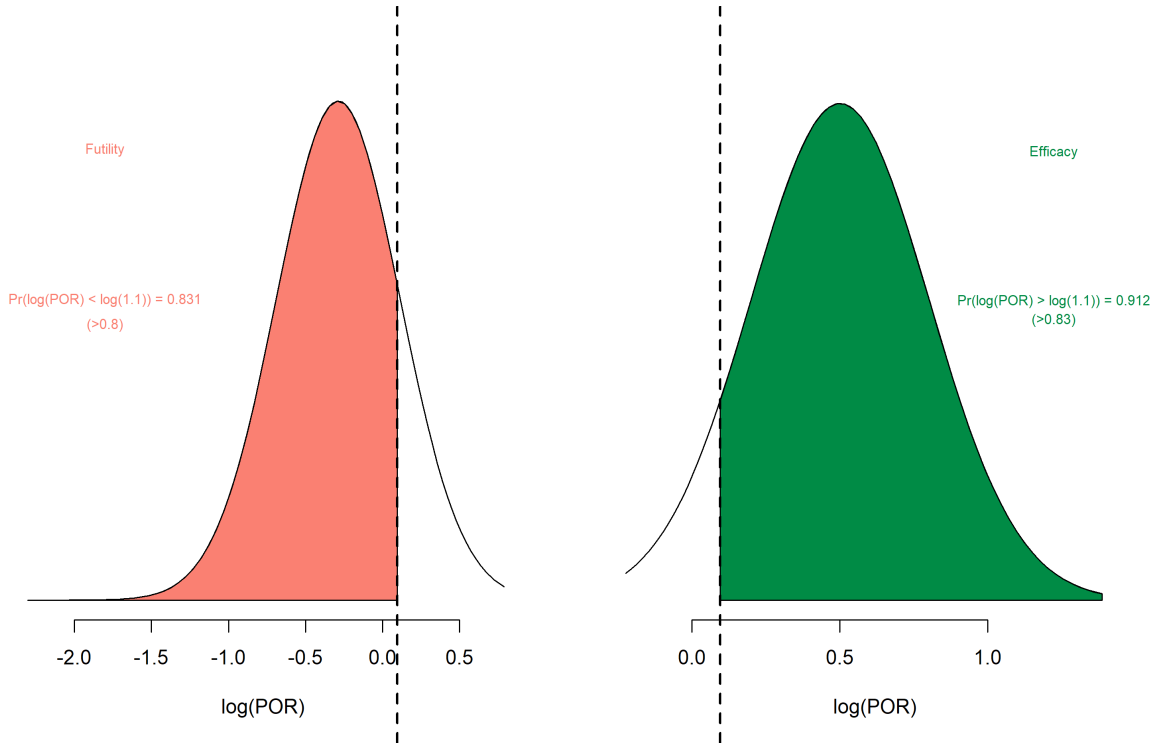

FIGURE 2: ILLUSTRATION OF STOPPING RULES – FUTILITY (LEFT PANEL) AND EFFICACY (RIGHT PANEL)

7 Statistical analysis of the primary endpoint

The primary endpoint will be analysed using Bayesian proportional odds logistic regression, with the effect of each investigational treatment represented as a proportional odds ratio (POR), with POR>1 relative to usual care indicating a favorable treatment. A Dirichlet distribution will be used as prior distribution for the $\boldsymbol{\alpha}$ threshold parameters / cut points in the proportional odds model. Minimally informative normal prior distributions will be used for the treatment effect terms in the model. The model will adjust for covariates used in the minimisation algorithm (these will not be included in the simulations). Analyses within each phenotype will be performed separately (i.e., stratified analyses by subtype). Each intervention arm will be compared to usual care through separate models (see figure 3 below).

The regression models will then be performed separately within each subphenotype:

- Anti-IL6 vs Usual care (red in the figure): $\text{logit}(P(Y \geq y))=\boldsymbol{\alpha}_{xHy}+\boldsymbol{\beta}_{xH}\,X_{xH}$

- Anti-TNF $\boldsymbol{\alpha}$ vs Usual care (purple in the figure): $\text{logit}(P(Y \geq y)) = \boldsymbol{\alpha}_{xAy} + \boldsymbol{\beta}_{xA} X_{xA}$

Where

- x is the subphenotype SRS 1 or SRS 2,
- $\boldsymbol{\alpha}_{xHy}$ and $\boldsymbol{\alpha}_{xAy}$ are the intercept for outcome level y in the subphenotype x,
- $\boldsymbol{\beta}_{xH}$ and $\boldsymbol{\beta}_{xA}$ are log(POR) which are the treatments effect for Anti-IL6 and Anti-TNF $\boldsymbol{\alpha}$ respectively in the subphenotype x,
- X_{xH} and X_{xA} are the treatment indicators (=1 if randomised to that treatment, 0 otherwise), in the subphenotype x.

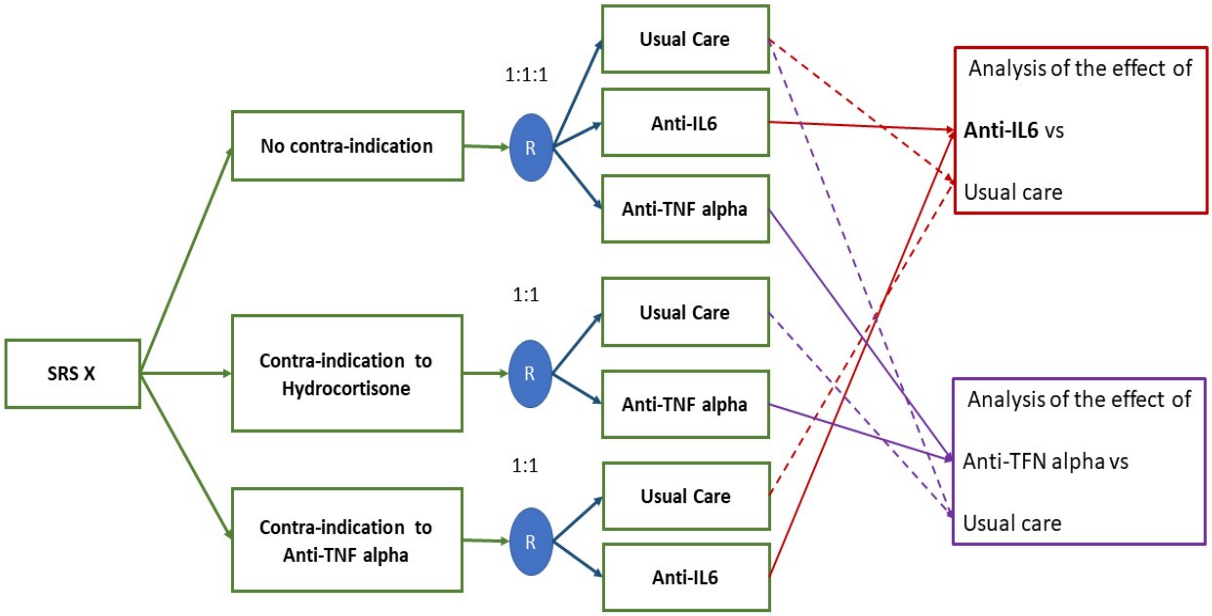

FIGURE 3: POPULATION INCLUDED IN THE ANALYSES

8 Simulations set up

8.1 Generating data

Data were generated independently within each subphenotype using the following process.
The first step was to generate participant **accrual** from a Poisson process.

1. For each month m, generate n times/durations between recruitment of participants (time between arrival of participants) u_i (i=1,..,n) from an exponential distribution with overall monthly recruitment rate equal to the number of sites recruiting that month * Number patients per site that month* split of subphenotype (see 8.5.2 for assumed recruitment rates).
2. The recruitment time for the participant i is then defined as $t_i = \sum_{j=1}^{i} u_j$.
3. Only keep those participants if t_i is less than or equals to 1 i.e patient included within the month m.
4. The recruitment time for participant i is then equal to $m+t_i$.

5. We allow for a data entry delay of 2 months so that the primary outcome is observed/ready for analysis at 2 months post randomisation (1 month to observe and 1 month to enter/clean data).

Then an **eligibility** random variable was simulated for each participant indicating whether the participant presents:

- No contraindications
- Contraindication to Anti-IL6
- Contraindication to Anti-TNF $\boldsymbol{\alpha}$

This was sampled with replacement with the probabilities 80%, 10%, 10% respectively.

The **treatment arm allocation** was then generated according to the (above) eligibility criteria:

- If the participant had no contraindication, they were assigned to one of the three treatment arms with a probability of 33%.
- If the participant had contraindication to Anti-IL6, they were assigned to usual care or Anti-TNF $\boldsymbol{\alpha}$ with a probability of 50%.
- If the participant had contraindication to Anti-TNF $\boldsymbol{\alpha}$, they were assigned to usual care or Anti-IL6 with a probability of 50%.

The **distribution of the primary endpoint** for each active treatment was calculated as follows, for an assumed POR value:

1. Calculate the cumulative probability of being in the category y or lower in the control group (see 8.5.1).
2. Calculate for each active treatment, the cumulative probability of being in the category y according to the POR value as follow:
    a. P(Y<=y)=exp(logit(P(Y<=y in the control group))+log(1/POR))/(1+exp(logit(P(Y<=y in the control group))+log(1/POR)))
    b. P(Y=y)=P(Y<=y)-P(Y<=y-1)

Then for each participant, the **primary endpoint was sampled** with replacement using the probabilities calculated as above according to their assigned arm.

8.2 Estimation, software and number of simulations

The Integrated Nested Laplace Approximation (INLA) technique was used to estimate the Bayesian models owing to the substantially faster runtime compared to Markov Chain Monte Carlo (MCMC) methods. Simulations have been run in R (R version 4.4.3; 2025-02-28 ucrt) using the R-INLA 24.12.11 package for Bayesian analyses. A limitation of this package is that ordinal outcomes for proportional odds logistic regression can have a maximum of 10 categories. The operating characteristic simulations are therefore based on a modified version of the primary outcome, which combines the original values into categories as shown the table below.

Organ support free days	New Mapping OSFD
-1	1
0	2
1-9	3
10-14	4
15-19	5
20-22	6
23, 24	7
25	8
26	9
27, 28	10

1000 trials were simulated for each design and scenario combination. We also simulated 10,000 trials under the global null scenario (for the non-binding futility design) to estimate the type I error more precisely, and under the "both work" scenario (POR = 1.37 for both arms in SRS1; POR = 1.42 for both arms in SRS2) and the "one works" scenario (POR1 = 1.37, POR2=1 for SRS1; POR1 = 1.42, POR2=1 for SRS2) so that power could be estimated more precisely (for the non-binding futility design).

8.3 Interim and final analyses

The first interim analysis was performed after a certain number of participants are due to complete follow up (**a minimum of 50 per intervention arm in a subtype**). The subsequent interim analyses were performed on a **6-monthly basis**. At the first interim analysis**, only** the **futility** stopping triggers may be applied (non-binding). At subsequent interim analyses, **both the efficacy and futility triggers** are evaluated.

The trial keeps on running until one of the early stopping rules is met for both intervention arms or the maximum sample size (or recruitment duration) is reached for the intervention arms in the subphenotype/endotype.

At each interim analysis there were patients that have completed follow up and enrolled patients that haven't completed follow up: the latter patients were not included in the interim analysis. Once an investigational arm has hit a statistical trigger (and thus recruitment to the intervention arm has ceased), we then follow up those patients that were randomised to the control or the investigational arm prior to the interim analysis that reached a statistical trigger, and perform a final analysis. Decisions can potentially be overturned once we've followed up all enrolled patients. Restarting of a closed intervention arm (due to overturning an interim decision at a final analysis) was not considered in these simulations as it would likely be infeasible in practice.

8.4 Optimal triggers

To determine the suitable values of the statistical triggers/thresholds POR_{eff}, p_{eff}, POR_{fut} and p_{fut}, simulations were carried out across a range of potential values (beginning with similar values to those used in PANTHER (https://panthertrial.org/). The other parameters used to assess the triggers were fixed at their main value (see 8.5).

8.4.1 Efficacy trigger

The two main key criteria used to assess and compare designs are:

- Type I error rate (one-sided) - should be minimised and should not exceed 10% for each treatment in each subphenotype (considered to be a reasonable limit for a phase II platform, as successful/promising treatments will be evaluated further at phase III). As the futility triggers are non-binding, a conservative estimate (or upper bound) of the type I error rate must assume that active interventions are never stopped for futility. The type I error rate is assessed when POR=1 (and using non-binding futility). No stopping rules were applied at the first interim analysis (when non-binding futility was assumed).
- Power - should be maximised and should be at least 70% for each treatment in either subphenotype for appropriate POR values. Again, we assume that active interventions are never stopped for futility.
- (Additionally we perform simulations assuming binding futility to assess the impact on power and sample size)

The ranges of values evaluated for each efficacy parameter were initially set as follows:

- Efficacy odds ratio threshold POR_{eff}: 1.075 to 1.15
- Efficacy probability threshold p_{eff}: 0.8 to 0.95.

Once the efficacy triggers were chosen, we then explored appropriate futility stopping rules.

Based on the results of several rounds of simulations, our chosen values are POR_{eff}=1.1 and p_{eff}=0.83.

8.4.2 Futility trigger

The futility triggers were assessed at the chosen efficacy triggers. The ranges of values evaluated for each futility parameter were initially set as follows:

- Futility odds ratio threshold POR_{fut}: 1.05 to 1.15
- Futility probability threshold p_{fut}: 0.7 to 0.8

Based on the results of several rounds of simulations (and clinician input), our chosen values are POR_{fut}=1.1 and p_{fut}=0.8.

8.5 Scenarios under chosen triggers

Certain design parameters were then varied one at the time with the other parameters fixed at their main value (see below). These sensitivity analyses are presented in the Supplementary Material.

8.5.1 Distribution of the primary endpoint

8.5.1.1 Main

For each SRS group, three data sources relating to OSFD in sepsis patients (observational data, VANISH and LeoPARDS trials) were averaged, and additionally, and a spline smoother were used for categories 1 to 26. The probabilities in the death category were a bit higher than expected, and based on secondary analysis results from the ADRENAL trial, we used a lower mortality rate in each SRS group (and redistributed the remaining probability equally amongst the other 29 categories).

category	Organ support free days	P control SRS 1	P control SRS 2
1	-1	0.2258	0.1757
2	0	0.0920	0.0622
3	1	0.0099	0.0076
4	2	0.0060	0.0096
5	3	0.0101	0.0060
6	4	0.0059	0.0101
7	5	0.0083	0.0087
8	6	0.0041	0.0116
9	7	0.0073	0.0144
10	8	0.0060	0.0096
11	9	0.0112	0.0054
12	10	0.0071	0.0054
13	11	0.0062	0.0160
14	12	0.0265	0.0067
15	13	0.0130	0.0094
16	14	0.0083	0.0103
17	15	0.0121	0.0257
18	16	0.0125	0.0122
19	17	0.0281	0.0212
20	18	0.0140	0.0396
21	19	0.0251	0.0190
22	20	0.0309	0.0386
23	21	0.0268	0.0343
24	22	0.0345	0.0407
25	23	0.0319	0.0406
26	24	0.0372	0.0517
27	25	0.0996	0.0904
28	26	0.1073	0.0874
29	27	0.0643	0.0773
30	28	0.0280	0.0526

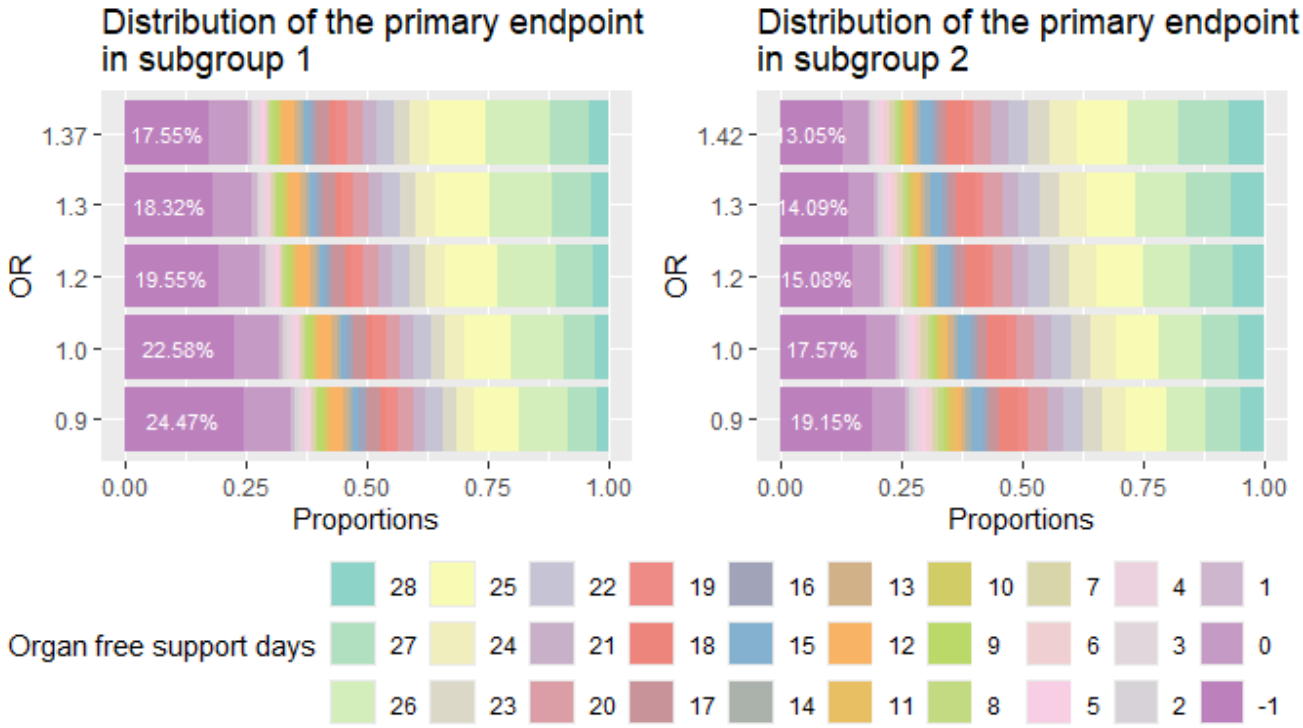

FIGURE 4: DISTRIBUTION OF THE PRIMARY ENDPOINT BY SRS GROUP FOR DIFFERENT ASSUMED TREATMENT EFFECT SIZES

8.5.1.2 Sensitivity analysis

The first sensitivity analysis assumed a higher mortality rate in the SRS1 subphenotype.

category	Organ support free days	P control SRS 1	P control SRS 2
1	-1	0.2758	0.1757
2	0	0.0903	0.0622
3	1	0.0082	0.0076
4	2	0.0043	0.0096
5	3	0.0084	0.0060
6	4	0.0042	0.0101
7	5	0.0066	0.0087
8	6	0.0024	0.0116
9	7	0.0056	0.0144
10	8	0.0043	0.0096
11	9	0.0095	0.0054
12	10	0.0054	0.0054
13	11	0.0045	0.0160
14	12	0.0248	0.0067
15	13	0.0113	0.0094
16	14	0.0066	0.0103
17	15	0.0104	0.0257
18	16	0.0108	0.0122
19	17	0.0264	0.0212
20	18	0.0123	0.0396
21	19	0.0234	0.0190
22	20	0.0292	0.0386
23	21	0.0251	0.0343

category	Organ support free days	P control SRS 1	P control SRS 2
24	22	0.0328	0.0407
25	23	0.0302	0.0406
26	24	0.0355	0.0517
27	25	0.0979	0.0904
28	26	0.1056	0.0874
29	27	0.0624	0.0773
30	28	0.0258	0.0527

The second sensitivity analysis assumed a higher mortality rate in the SRS2 subphenotype.

category	Organ support free days	P control SRS 1	P control SRS 2
1	-1	0.2258	0.2257
2	0	0.0920	0.0605
3	1	0.0099	0.0059
4	2	0.0060	0.0079
5	3	0.0101	0.0043
6	4	0.0059	0.0084
7	5	0.0083	0.0070
8	6	0.0041	0.0099
9	7	0.0073	0.0127
10	8	0.0060	0.0079
11	9	0.0112	0.0037
12	10	0.0071	0.0037
13	11	0.0062	0.0143
14	12	0.0265	0.0050
15	13	0.0130	0.0077
16	14	0.0083	0.0086
17	15	0.0121	0.0240
18	16	0.0125	0.0105
19	17	0.0281	0.0195
20	18	0.0140	0.0379
21	19	0.0251	0.0173
22	20	0.0309	0.0369
23	21	0.0268	0.0326
24	22	0.0345	0.0390
25	23	0.0319	0.0389
26	24	0.0372	0.0500
27	25	0.0996	0.0887
28	26	0.1073	0.0857
29	27	0.0643	0.0756
30	28	0.0280	0.0509

In the third sensitivity analysis, the tail top probabilities (-1, 0 categories) are kept the same as in the main analysis and the probabilities of the other categories (1-28) are evenly distributed.

category	Organ support free days	P control SRS 1	P control SRS 2
1	-1	0.2258	0.1757
2	0	0.0920	0.0622
3	1	0.0244	0.0272
4	2	0.0244	0.0272
5	3	0.0244	0.0272
6	4	0.0244	0.0272
7	5	0.0244	0.0272
8	6	0.0244	0.0272
9	7	0.0244	0.0272
10	8	0.0244	0.0272
11	9	0.0244	0.0272
12	10	0.0244	0.0272
13	11	0.0244	0.0272
14	12	0.0244	0.0272
15	13	0.0244	0.0272
16	14	0.0244	0.0272
17	15	0.0244	0.0272
18	16	0.0244	0.0272
19	17	0.0244	0.0272
20	18	0.0244	0.0272
21	19	0.0244	0.0272
22	20	0.0244	0.0272
23	21	0.0244	0.0272
24	22	0.0244	0.0272
25	23	0.0244	0.0272
26	24	0.0244	0.0272
27	25	0.0244	0.0272
28	26	0.0242	0.0272
29	27	0.0240	0.0274
30	28	0.0240	0.0275

8.5.2 Recruitment rate

8.5.2.1 Main

The cumulative number of patients included in each subphenotype was determined by multiplying the total number of patients per month by the proportion of patients in each subphenotype (split between subphenotype see 8.5.3). We assume a total of 60 sites with a recruitment rate of 1.5 patients per site (throughout the trial), and that it takes 21 months to get all 60 sites up and running. The UK will contribute 40 sites and Australia will contribute 20 sites.

Month	Number of sites	Number patients per site	Number of patients per month total	Cumulative number of patients
1	1	1.5	1.5	1.5
2	2	1.5	3.0	4.5
3	3	1.5	4.5	9.0
4	5	1.5	7.5	16.5
5	5	1.5	7.5	24.0
6	10	1.5	15.0	39.0
7	10	1.5	15.0	54.0
8	15	1.5	22.5	76.5
9	15	1.5	22.5	99.0
10	20	1.5	30.0	129.0
11	20	1.5	30.0	159.0
12	25	1.5	37.5	196.5
13	25	1.5	37.5	234.0
14	30	1.5	45.0	279.0
15	30	1.5	45.0	324.0
16	35	1.5	52.5	376.5
17	40	1.5	60.0	436.5
18	45	1.5	67.5	504.0
19	50	1.5	75.0	579.0
20	55	1.5	82.5	661.5
21	60	1.5	90.0	751.5
22	60	1.5	90.0	841.5
23	60	1.5	90.0	931.5
24	60	1.5	90.0	1,021.5
25	60	1.5	90.0	1,111.5
26	60	1.5	90.0	1,201.5
27	60	1.5	90.0	1,291.5
28	60	1.5	90.0	1,381.5
29	60	1.5	90.0	1,471.5
30	60	1.5	90.0	1,561.5
31	60	1.5	90.0	1,651.5
32	60	1.5	90.0	1,741.5
33	60	1.5	90.0	1,831.5
34	60	1.5	90.0	1,921.5
35	60	1.5	90.0	2,011.5
36	60	1.5	90.0	2,101.5
37	60	1.5	90.0	2,191.5
38	60	1.5	90.0	2,281.5
39	60	1.5	90.0	2,371.5
40	60	1.5	90.0	2,461.5
41	60	1.5	90.0	2,551.5
42	60	1.5	90.0	2,641.5

8.5.2.2 Sensitivity analysis

In the first sensitivity analysis, we assumed a smaller number of patients per site.

Month	Number of sites	Number patients per site	Number of patients per month total	Cumulative number of patients
1	1	1.25	1.25	1.25
2	2	1.25	2.50	3.75
3	3	1.25	3.75	7.50
4	5	1.25	6.25	13.75
5	5	1.25	6.25	20.00
6	10	1.25	12.50	32.50
7	10	1.25	12.50	45.00
8	15	1.25	18.75	63.75
9	15	1.25	18.75	82.50
10	20	1.25	25.00	107.50
11	20	1.25	25.00	132.50
12	25	1.25	31.25	163.75
13	25	1.25	31.25	195.00
14	30	1.25	37.50	232.50
15	30	1.25	37.50	270.00
16	35	1.25	43.75	313.75
17	40	1.25	50.00	363.75
18	45	1.25	56.25	420.00
19	50	1.25	62.50	482.50
20	55	1.25	68.75	551.25
21	60	1.25	75.00	626.25
22	60	1.25	75.00	701.25
23	60	1.25	75.00	776.25
24	60	1.25	75.00	851.25
25	60	1.25	75.00	926.25
26	60	1.25	75.00	1,001.25
27	60	1.25	75.00	1,076.25
28	60	1.25	75.00	1,151.25
29	60	1.25	75.00	1,226.25
30	60	1.25	75.00	1,301.25
31	60	1.25	75.00	1,376.25
32	60	1.25	75.00	1,451.25
33	60	1.25	75.00	1,526.25
34	60	1.25	75.00	1,601.25
35	60	1.25	75.00	1,676.25
36	60	1.25	75.00	1,751.25
37	60	1.25	75.00	1,826.25
38	60	1.25	75.00	1,901.25
39	60	1.25	75.00	1,976.25
40	60	1.25	75.00	2,051.25
41	60	1.25	75.00	2,126.25
42	60	1.25	75.00	2,201.25

In the second sensitivity analysis, we assumed a slower initial recruitment rate (fewer per site, and that it took longer to reach 60 sites).

Month	Number of sites	Number patients per site	Number of patients per month total	Cumulative number of patients
1	1	1.00	1.00	1.00
2	2	1.00	2.00	3.00
3	3	1.00	3.00	6.00
4	4	1.00	4.00	10.00
5	5	1.00	5.00	15.00
6	6	1.00	6.00	21.00
7	8	1.25	10.00	31.00
8	10	1.25	12.50	43.50
9	12	1.25	15.00	58.50
10	15	1.25	18.75	77.25
11	18	1.25	22.50	99.75
12	21	1.25	26.25	126.00
13	25	1.50	37.50	163.50
14	28	1.50	42.00	205.50
15	31	1.50	46.50	252.00
16	34	1.50	51.00	303.00
17	37	1.50	55.50	358.50
18	40	1.50	60.00	418.50
19	43	1.50	64.50	483.00
20	46	1.50	69.00	552.00
21	49	1.50	73.50	625.50
22	52	1.50	78.00	703.50
23	55	1.50	82.50	786.00
24	58	1.50	87.00	873.00
25	60	1.50	90.00	963.00
26	60	1.50	90.00	1,053.00
27	60	1.50	90.00	1,143.00
28	60	1.50	90.00	1,233.00
29	60	1.50	90.00	1,323.00
30	60	1.50	90.00	1,413.00
31	60	1.50	90.00	1,503.00
32	60	1.50	90.00	1,593.00
33	60	1.50	90.00	1,683.00
34	60	1.50	90.00	1,773.00
35	60	1.50	90.00	1,863.00
36	60	1.50	90.00	1,953.00
37	60	1.50	90.00	2,043.00
38	60	1.50	90.00	2,133.00
39	60	1.50	90.00	2,223.00
40	60	1.50	90.00	2,313.00
41	60	1.50	90.00	2,403.00
42	60	1.50	90.00	2,493.00

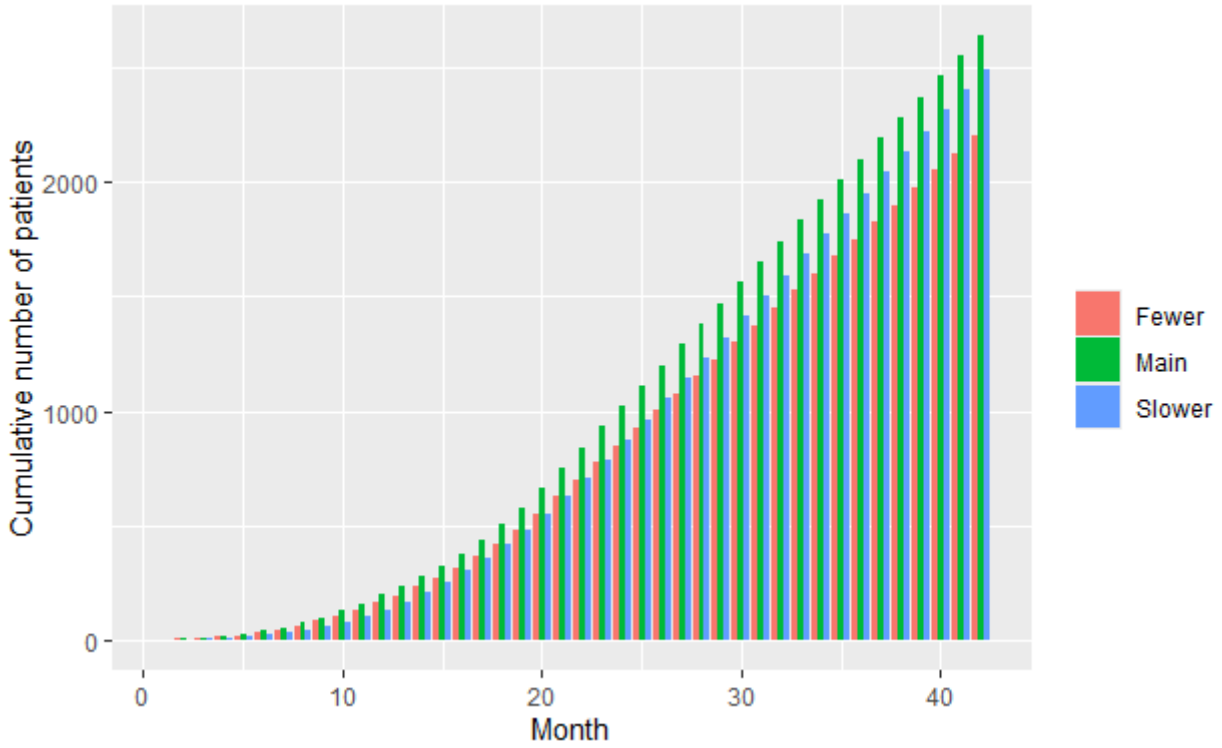

FIGURE 5: CUMULATIVE NUMBER OF PATIENTS RECRUITED ACROSS DIFFERENT SCENARIOS

8.5.3 Other parameters

Parameter	Main	Sensitivity analysis
Treatment effect size (POR for Anti-IL6, POR for anti-TNF alpha)	(1.37,1.37) (SRS 1; 5% absolute decrease in mortality; "power" scenario) (1.42,1.42) (SRS 2; 4.5% absolute decrease in mortality; "power" scenario) (1,1) (type I error)	POR for treatment 1 = power OR POR for treatment 2: 0.9, 1, 1.2, 1.3, 1.4
Split between phenotype	40% (SRS 1) vs 60% (SRS 2)	30% (SRS 1) vs 70% (SRS 2) 50% (SRS 1) vs 50% (SRS 2)

Alternative interim analysis schedules were also explored where we varied the timing of the first interim analysis, the decisions that could be made at the first interim analyses, and the timing of subsequent interim analyses (results not presented here).

8.6 Assumptions

We assumed:

- Drop outs/missing data: 3% of dropouts, simulated from a binomial distribution. Minimal withdrawals are expected owing to the nature of the condition being treated (participants will require constant support in ICU).
- Data delays: it may take up to 28 days to observe each participant's primary outcome, and it may take a further month for the data to be entered into the database and ready for analysis. Thus, we assumed that participants' data would not be ready for analysis until 2 months after their randomisation date, and may not be observed at the time of the interim analysis. We have

excluded participants that won't have reached 28 days follow up (+ 1 month, i.e. 2 months post-randomisation date) in the interim analyses.
- No information borrowing is used across the treatment effects for SRS1 and SRS2 as we assume that the effects could be quite different between subphenotypes.
- A total of 60 recruiting sites (40 in the UK and 20 in Australia) are part of the trial,
- Although the real-world platform will adjust for covariates used in the minimisation algorithm in the analyses, no adjustment for covariates in the regression model is used in the simulations.
- Two active treatments arms and one control arm in each subphenotype.
- Fixed randomisation was used (with arm dropping). We used equal randomisation for the simulations. The randomisation was stratified on the eligibility criteria.

8.7 Performance measures

In order to compare the different scenarios, different measures were calculated:
- Probability that the efficacy criteria is met at the final analysis.
- Probability that the futility criteria is met at the final analysis.
- Probability of early stopping for efficacy.
- Probability of early stopping for futility.
- Type I error percentage in each subphenotype when the POR=1.
- Mean and quantiles of cumulative sample size required within each subphenotype.
- Mean and quantiles of time the first interim analysis was performed.
- Mean and quantiles of number of interim analyses performed within each subphenotype and active treatment arm.

9 Simulation results

9.1 Main results

9.1.1 Non-binding futility

9.1.1.1 Type I and power

The following table presents results of the type I error (probability of declaring superiority while TRUE POR = 1) and power (probability to declare superiority while TRUE POR > 1) at the final analysis. (POR stands for proportional odds ratio; POR 1 is the POR for intervention/treatment 1, and POR 2 is the POR for intervention/treatment 2).

The results presented below assume POR = 1.37 for treatment 1 for SRS1 and POR = 1.42 for treatment 1 for SRS2 for the power scenarios, and the values for treatment 2 vary (POR 2). The first half of the table presents the power / type I error for treatment 1, and the second half of the table presents the power / type I error for treatment 2.

TABLE 1: MAIN RESULTS: POWER AND TYPE I ERROR (ASSUMING NON-BINDING FUTILITY)

Group	TRUE POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Subgroup 1		Subgroup 2	
						No Superiority [1]	Superiority [1]	No Superiority [1]	Superiority [1]
Treatment1									
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	889/ 990 (90%)	**101/ 990 (10%)**	889/ 976 (91%)	**87/ 976 (8.9%)**
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	269/ 985 (27%)	716/ 985 (73%)	153/ 974 (16%)	821/ 974 (84%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	262/ 991 (26%)	729/ 991 (74%)	158/ 971 (16%)	813/ 971 (84%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	246/ 989 (25%)	743/ 989 (75%)	151/ 967 (16%)	816/ 967 (84%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	233/ 990 (24%)	**757/ 990 (76%)**		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	234/ 988 (24%)	754/ 988 (76%)		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	270/ 987 (27%)	717/ 987 (73%)	155/ 970 (16%)	815/ 970 (84%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6			149/ 970 (15%)	**821/ 970 (85%)**
Treatment2									
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	874/ 990 (88%)	**116/ 990 (12%)**	883/ 976 (90%)	**93/ 976 (9.5%)**
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	952/ 985 (97%)	33/ 985 (3.4%)	958/ 974 (98%)	16/ 974 (1.6%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	518/ 991 (52%)	473/ 991 (48%)	577/ 971 (59%)	394/ 971 (41%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	320/ 989 (32%)	669/ 989 (68%)	354/ 967 (37%)	613/ 967 (63%)

Group	TRUE POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Subgroup 1		Subgroup 2	
						No Superiority [1]	Superiority [1]	No Superiority [1]	Superiority [1]
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	215/ 990 (22%)	**775/ 990 (78%)**		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	183/ 988 (19%)	805/ 988 (81%)		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	881/ 987 (89%)	106/ 987 (11%)	887/ 970 (91%)	83/ 970 (8.6%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6			162/ 970 (17%)	**808/ 970 (83%)**
[1]n/ N (%)									

The one-sided type I error was approximately 10% for each comparison of the intervention arms to the control (pairwise type I error) for each SRS endotype. Note that when 10,000 simulations were performed for the global null scenario, one-sided type I errors of 10-11% in SRS 1 and 9.2% for SRS 2 (Table 2). Assuming a POR = 1.37 for an intervention arm in SRS1, the power was >73%, and assuming a POR = 1.42 for an intervention arm in SRS2 the power was ≥ 83%. There was a low probability of declaring superiority when a harmful treatment effect (POR=0.9) was assumed (3.4% in SRS1 and 1.6% in SRS 2).

TABLE 2: MAIN RESULTS: POWER AND TYPE I ERROR 10000 SIMULATIONS (ASSUMING NON-BINDING FUTILITY)

Group	TRUE POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Subgroup 1		Subgroup 2	
						No Superiority [1]	Superiority [1]	No Superiority [1]	Superiority [1]
Treatment1									
	scenario								
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	8,871/ 9,899 (90%)	**1,028/ 9,899 (10%)**	8,877/ 9,772 (91%)	**895/ 9,772 (9.2%)**
Treatment2									
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	8,811/ 9,899 (89%)	**1,088/ 9,899 (11%)**	8,873/ 9,772 (91%)	**899/ 9,772 (9.2%)**
Treatment1									

Group	TRUE POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Subgroup 1		Subgroup 2	
						No Superiority [1]	Superiority [1]	No Superiority [1]	Superiority [1]
	POR = 1.37, 1.37 (SRS1); POR = 1.42, 1.42 (SRS2)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	2255/9752 (23%)	**7497/9752 (77%)**	1547/9607 (16%)	**8060/9607 (84%)**
Treatment2									
	POR = 1.37, 1.37 (SRS1); POR = 1.42, 1.42 (SRS2)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	2207/9752 (23%)	**7545/9752 (77%)**	1503/9607 (16%)	**8104/9607 (84%)**

[1]n/ N (%)

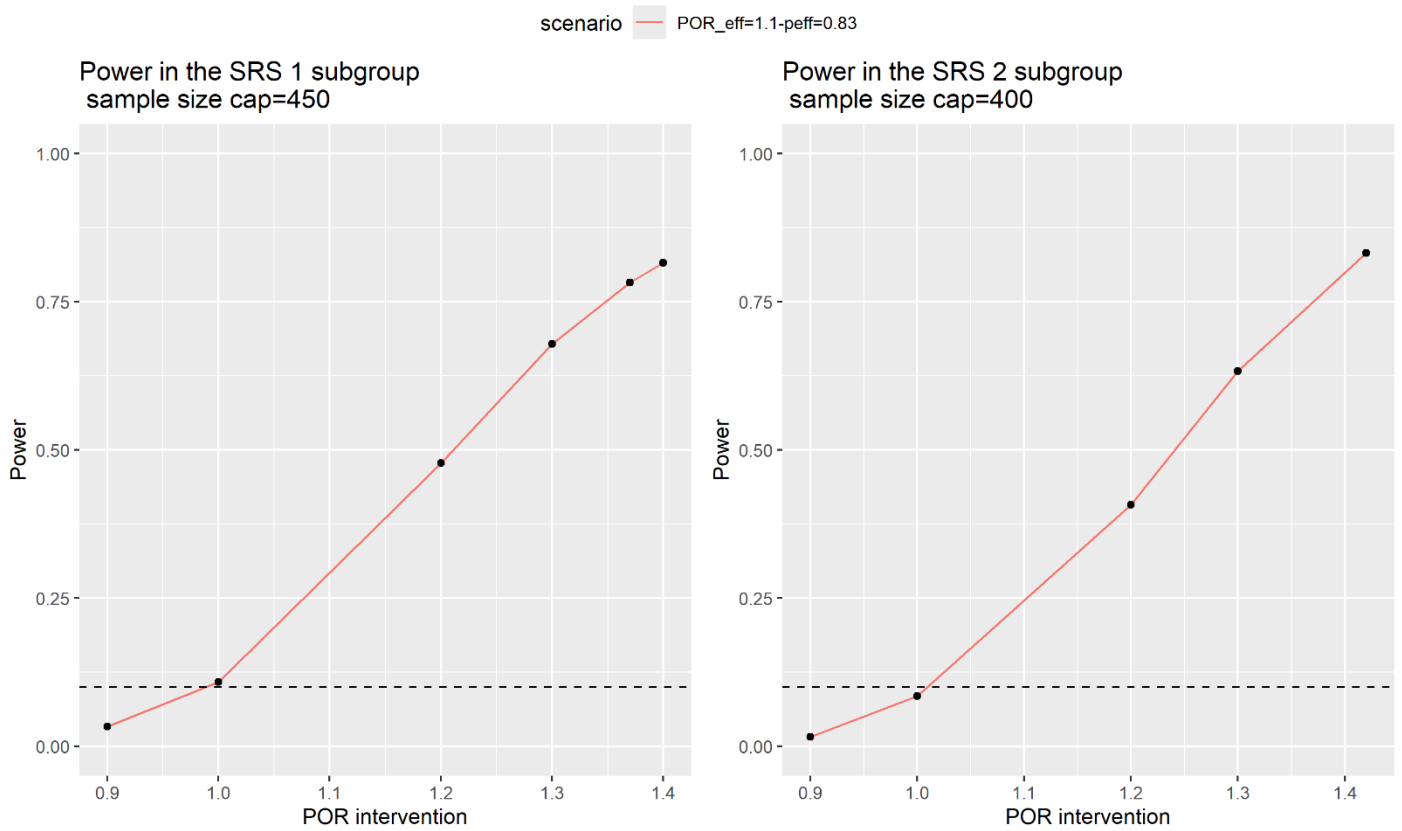

FIGURE 6: MAIN RESULTS: TREATMENT EFFECT SIZE (POR) VS POWER (PAIRWISE) FOR SRS 1 (LEFT PANEL) AND SRS 2 (RIGHT PANEL) (ASSUMING NON-BINDING FUTILITY). DOTTED HORIZONTAL LINES ARE PLOTTED AT 0.1 (TYPE I ERROR THRESHOLD).

9.1.1.2 Stopping reason and Sample size

The following table presents the reason why the trial stopped and the sample size (for an intervention arm) for each scenario.

TABLE 3: MAIN RESULTS: STOPPING REASON AND SAMPLE SIZE (ASSUMING NON-BINDING FUTILITY)

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 990[1]	SRS 2 N = 976[1]
Treatment1								
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	811/ 990 (82%)	
						Reached sample size cap	33/ 990 (3.3%)	852/ 976 (87%)
						Superiority at interim analysis	146/ 990 (15%)	124/ 976 (13%)
						Sample size	319 (73) 333 [324-343] (121-450)	376 (69) 400 [400-400] (137-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	236/ 985 (24%)	
						Reached sample size cap		126/ 974 (13%)
						Superiority at interim analysis	749/ 985 (76%)	848/ 974 (87%)
						Sample size	208 (85) 184 [135-308] (114-402)	229 (95) 164 [154-277] (134-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	227/ 991 (23%)	
						Reached sample size cap	15/ 991 (1.5%)	144/ 971 (15%)
						Superiority at interim analysis	749/ 991 (76%)	827/ 971 (85%)
						Sample size	216 (96) 186 [135-322] (114-450)	231 (97) 164 [154-318] (134-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	199/ 989 (20%)	

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 990[1]	SRS 2 N = 976[1]
						Reached sample size cap	25/ 989 (2.5%)	149/ 967 (15%)
						Superiority at interim analysis	765/ 989 (77%)	818/ 967 (85%)
						Sample size	220 (101) 189 [135-323] (114-450)	232 (98) 164 [154-320] (134-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	182/ 990 (18%)	
						Reached sample size cap	35/ 990 (3.5%)	
						Superiority at interim analysis	773/ 990 (78%)	
						Sample size	223 (106) 190 [135-327] (114-450)	
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	181/ 988 (18%)	
						Reached sample size cap	34/ 988 (3.4%)	
						Superiority at interim analysis	773/ 988 (78%)	
						Sample size	224 (107) 185 [135-328] (114-450)	
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	234/ 987 (24%)	
						Reached sample size cap	5/ 987 (0.5%)	129/ 970 (13%)
						Superiority at interim analysis	748/ 987 (76%)	841/ 970 (87%)
						Sample size	209 (88) 174 [135-317] (114-450)	229 (95) 164 [154-282] (134-400)

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 990[1]	SRS 2 N = 976[1]
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Reached sample size cap		168/ 970 (17%)
						Superiority at interim analysis		802/ 970 (83%)
						Sample size		235 (100) 165 [154-326] (134-400)
Treatment2								
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	799/ 990 (81%)	
						Reached sample size cap	31/ 990 (3.1%)	843/ 976 (86%)
						Superiority at interim analysis	160/ 990 (16%)	133/ 976 (14%)
						Sample size	317 (73) 333 [322-343](120-450)	372 (75) 400 [400-400](137-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	706/ 985 (72%)	
						Reached sample size cap	223/ 985 (23%)	940/ 974 (97%)
						Superiority at interim analysis	56/ 985 (5.7%)	34/ 974 (3.5%)
						Sample size	390 (73) 416 [345-447](118-450)	393 (41) 400 [400-400](138-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	387/ 991 (39%)	
						Reached sample size cap	91/ 991 (9.2%)	532/ 971 (55%)
						Superiority at interim analysis	513/ 991 (52%)	439/ 971 (45%)

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 990[1]	SRS 2 N = 976[1]
						Sample size	287 (119) 326 [140-393](118-450)	313 (108) 400 [164-400](137-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	233/ 989 (24%)	
						Reached sample size cap	52/ 989 (5.3%)	351/ 967 (36%)
						Superiority at interim analysis	704/ 989 (71%)	616/ 967 (64%)
						Sample size	244 (112) 212 [136-340](116-450)	278 (109) 298 [158-400](137-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	166/ 990 (17%)	
						Reached sample size cap	28/ 990 (2.8%)	
						Superiority at interim analysis	796/ 990 (80%)	
						Sample size	217 (103) 145 [134-310](116-450)	
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	139/ 988 (14%)	
						Reached sample size cap	25/ 988 (2.5%)	
						Superiority at interim analysis	824/ 988 (83%)	
						Sample size	208 (99) 142 [134-270](115-450)	
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Max recruitment period	641/ 987 (65%)	

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 990[1]	SRS 2 N = 976[1]
						Reached sample size cap	193/ 987 (20%)	850/ 970 (88%)
						Superiority at interim analysis	153/ 987 (16%)	120/ 970 (12%)
						Sample size	367 (95) 399 [334-444](120-450)	373 (75) 400 [400-400](137-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Stopping reason		
						Reached sample size cap		189/ 970 (19%)
						Superiority at interim analysis		781/ 970 (81%)
						Sample size		239 (101) 166 [154-331](137-400)

[1]n/ N (%); Mean (SD) Median [Q1-Q3] (Min-Max)

TABLE 4: MAIN RESULTS: SAMPLE SIZE (ASSUMING NON-BINDING FUTILITY)

Group	Sample size cap	TRUE POR (POR 1,POR 2)	Efficacy stopping rule (POR eff - peff)	Futility stopping rule (POR fut - pfut)	Triggers not implemented at first interim analysis	SRS1 vs SRS2	total N = 999[1]	Treatment 1 N = 999[1]	Treatment 2 N = 999[1]
SRS 1									
	450	1.37,0.9	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	1,021 (130) 1,049 [1,032-1,063] [376-1,160]	208 (85) 184 [135-308] [114-402]	390 (73) 416 [345-447] [118-450]
	450	1.37,1.2	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	843 (262) 1,017 [626-1,054] [376-1,160]	216 (96) 186 [135-322] [114-450]	287 (119) 326 [140-393] [118-450]
	450	1.37,1.3	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	773 (268) 840 [442-1,047] [368-1,160]	220 (101) 189 [135-323] [114-450]	244 (112) 212 [136-340] [116-450]
	450	1.37,1.37	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	735 (267) 658 [436-1,038] [368-1,160]	223 (106) 190 [135-327] [114-450]	217 (103) 145 [134-310] [116-450]
	450	1.37,1.4	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	719 (266) 648 [434-1,034] [368-1,160]	224 (107) 185 [135-328] [114-450]	208 (99) 142 [134-270] [115-450]
	450	1.37,1	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	982 (185) 1,046 [1,022-1,062] [385-1,160]	209 (88) 174 [135-317] [114-450]	367 (95) 399 [334-444] [120-450]
	450	1,1	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	1,025 (126) 1,052 [1,031-1,072] [396-1,160]	319 (73) 333 [324-343] [121-450]	317 (73) 333 [322-343] [120-450]
SRS 2									
	400	1.42,0.9	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	1,052 (144) 992 [978-1,123] [439-1,282]	229 (95) 164 [154-277] [134-400]	393 (41) 400 [400-400] [138-400]
	400	1.42,1.2	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	908 (260) 977 [784-1,101] [432-1,282]	231 (97) 164 [154-318] [134-400]	313 (108) 400 [164-400] [137-400]
	400	1.42,1.3	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	847 (268) 852 [514-1,084] [432-1,282]	232 (98) 164 [154-320] [134-400]	278 (109) 298 [158-400] [137-400]
	400	1.42,1.42	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	784 (265) 809 [498-986] [432-1,281]	235 (100) 165 [154-326] [134-400]	239 (101) 166 [154-331] [137-400]
	400	1.42,1	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	1,014 (201) 989 [975-1,119] [432-1,282]	229 (95) 164 [154-282] [134-400]	373 (75) 400 [400-400] [137-400]
	400	1,1	1.1 - 0.83	No futility rules	TRUE	0.4 vs 0.6	1,198 (159) 1,265 [1,258-1,269] [439-1,287]	376 (69) 400 [400-400] [137-400]	372 (75) 400 [400-400] [137-400]

[1]Mean (SD) Median [Q1-Q3] [Min-Max]

9.1.1.3 Timing of the trial analyses

TABLE 5: MAIN RESULTS: TIME OF FIRST INTERIM ANALYSIS (MONTHS) AND TIME OF THE FINAL ANALYSIS (MONTHS) (ASSUMING NON-BINDING FUTILITY)

POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS 1 SRS1 vs SRS2	Analysis	SRS 1 N = 990[1]	SRS 2 N = 976[1]
POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.74 (0.50) 15.72 [15.39-16.08] (14.13-17.66)
POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	41.6 (5.9) 44.0 [44.0-44.0] (23.6-44.0)	36.6 (4.7) 38.0 [38.0-39.0] (21.7-41.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.73 (0.51) 15.72 [15.38-16.08] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	32.1 (7.7) 30.1 [25.5-42.1] (23.6-44.0)	27.2 (5.9) 23.3 [22.6-29.5] (21.1-40.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.73 (0.51) 15.72 [15.38-16.09] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	32.1 (7.8) 30.2 [25.5-42.7] (23.6-44.0)	27.0 (5.7) 23.3 [22.6-29.4] (21.1-40.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.45 [18.06-18.76] (16.55-20.05)	15.73 (0.50) 15.72 [15.38-16.08] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	32.0 (7.6) 30.2 [25.5-37.8] (23.6-44.0)	26.9 (5.5) 23.3 [22.6-29.3] (21.1-40.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.45 [18.07-18.76] (16.55-20.05)	
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	31.9 (7.5) 30.3 [25.5-37.8] (23.6-44.0)	

POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS 1 SRS1 vs SRS2	Analysis	SRS 1 N = 990[1]	SRS 2 N = 976[1]
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.45 [18.07-18.76] (16.55-20.05)	
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	31.8 (7.5) 30.1 [25.5-37.8] (23.6-44.0)	
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.74 (0.51) 15.72 [15.39-16.09] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	32.1 (7.8) 29.9 [25.5-42.6] (23.6-44.0)	27.2 (5.9) 23.3 [22.7-29.5] (21.1-40.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time first interim analysis (months)		15.73 (0.51) 15.72 [15.38-16.08] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	No futility rules	TRUE	0.4 vs 0.6	Time final analysis for treatment 1 (months)		26.8 (5.3) 23.3 [22.6-29.3] (21.1-40.0)

[1]Mean (SD) Median [Q1-Q3] (Min-Max)

9.1.2 Binding futility

9.1.2.1 Type I and power

The following table presents results of the type I error (probability to reach superiority while TRUE POR = 1) and power (probability to reach superiority while TRUPOR E > 1) at the final analysis with binding futility.

TABLE 6: MAIN RESULTS: POWER AND TYPE I ERROR (ASSUMING BINDING FUTILITY)

						Subgroup 1		Subgroup 2	
Group	TRUE POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	No Superiority [1]	Superiority [1]	No Superiority [1]	Superiority [1]
Treatment1									
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	899/ 994 (90%)	**95/ 994 (9.6%)**	896/ 972 (92%)	**76/ 972 (7.8%)**
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	255/ 989 (26%)	734/ 989 (74%)	179/ 974 (18%)	795/ 974 (82%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	286/ 990 (29%)	704/ 990 (71%)	198/ 969 (20%)	771/ 969 (80%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	280/ 989 (28%)	709/ 989 (72%)	193/ 968 (20%)	775/ 968 (80%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	279/ 990 (28%)	**711/ 990 (72%)**		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	276/ 989 (28%)	713/ 989 (72%)		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	271/ 986 (27%)	715/ 986 (73%)	181/ 967 (19%)	786/ 967 (81%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6			197/ 971 (20%)	**774/ 971 (80%)**
Treatment2									
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	891/ 994 (90%)	**103/ 994 (10%)**	890/ 972 (92%)	**82/ 972 (8.4%)**

Group	TRUE POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Subgroup 1		Subgroup 2	
						No Superiority [1]	Superiority [1]	No Superiority [1]	Superiority [1]
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	958/ 989 (97%)	31/ 989 (3.1%)	958/ 974 (98%)	16/ 974 (1.6%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	544/ 990 (55%)	446/ 990 (45%)	612/ 969 (63%)	357/ 969 (37%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	357/ 989 (36%)	632/ 989 (64%)	393/ 968 (41%)	575/ 968 (59%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	254/ 990 (26%)	**736/ 990 (74%)**		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	213/ 989 (22%)	776/ 989 (78%)		
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	890/ 986 (90%)	96/ 986 (9.7%)	888/ 967 (92%)	79/ 967 (8.2%)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6			209/ 971 (22%)	**762/ 971 (78%)**

[1]n/ N (%)

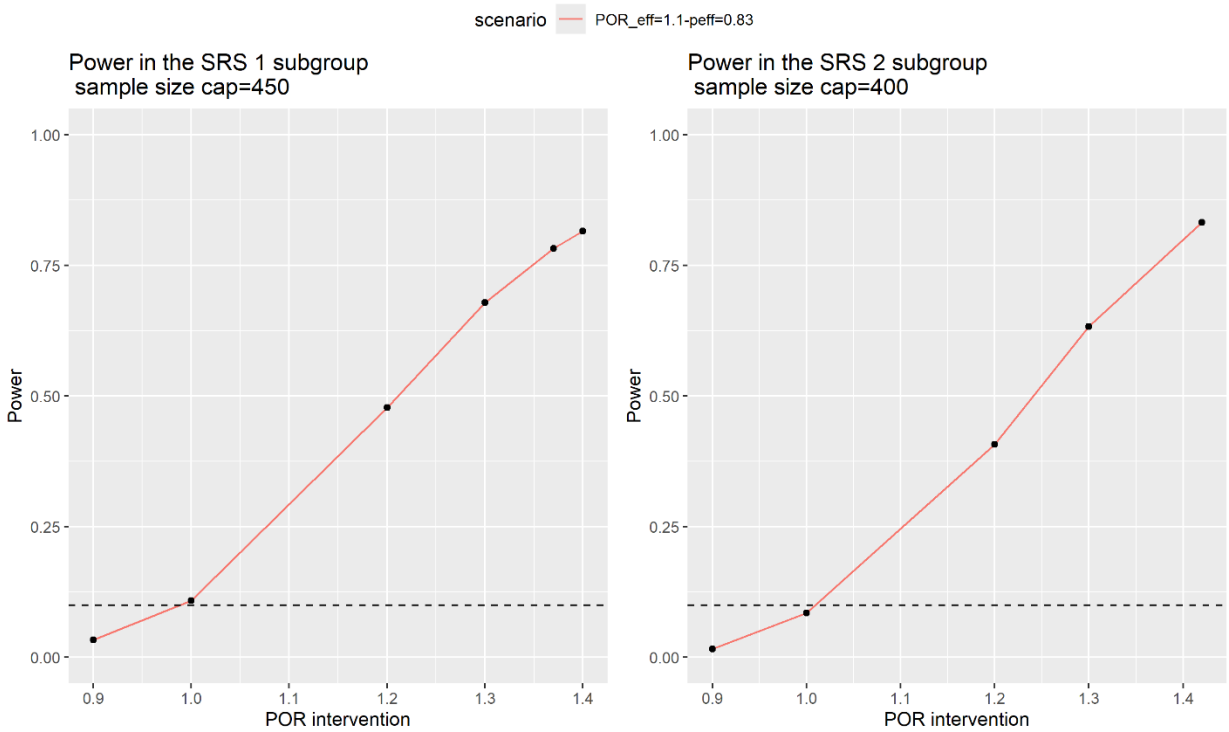

FIGURE 7: MAIN RESULTS: TREATMENT EFFECT SIZE (POR) VS POWER (PAIRWISE) FOR SRS 1 (LEFT PANEL) AND SRS 2 (RIGHT PANEL) (ASSUMING BINDING FUTILITY). DOTTED HORIZONTAL LINES ARE PLOTTED AT 0.1 (TYPE I ERROR THRESHOLD)

9.1.2.2 Stopping reason and Sample size

The following table presents reason why the trial stopped and the sample size (for each intervention) for each scenario.

TABLE 7: MAIN RESULTS: STOPPING REASON AND SAMPLE SIZE (ASSUMING BINDING FUTILITY)

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 994[1]	SRS 2 N = 972[1]
Treatment1								
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	582/ 994 (59%)	560/ 972 (58%)
						Max recruitment period	187/ 994 (19%)	
						Reached sample size cap	79/ 994 (7.9%)	300/ 972 (31%)
						Superiority at interim analysis	146/ 994 (15%)	112/ 972 (12%)
						Sample size	211 (133) 170 [72-334] (58-450)	236 (138) 212 [73-400] (61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	107/ 989 (11%)	69/ 974 (7.1%)
						Max recruitment period	52/ 989 (5.3%)	
						Reached sample size cap	62/ 989 (6.3%)	123/ 974 (13%)
						Superiority at interim analysis	768/ 989 (78%)	782/ 974 (80%)
						Sample size	209 (108) 169 [134-275] (58-450)	228 (100) 198 [155-319] (61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	112/ 990 (11%)	66/ 969 (6.8%)
						Max recruitment period	102/ 990 (10%)	
						Reached sample size cap	35/ 990 (3.5%)	116/ 969 (12%)
						Superiority at interim analysis	741/ 990 (75%)	787/ 969 (81%)

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 994[1]	SRS 2 N = 972[1]
						Sample size	196 (101) 141 [132-251] (58-450)	220 (99) 164 [153-297] (61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	109/ 989 (11%)	69/ 968 (7.1%)
						Max recruitment period	106/ 989 (11%)	
						Reached sample size cap	35/ 989 (3.5%)	115/ 968 (12%)
						Superiority at interim analysis	739/ 989 (75%)	784/ 968 (81%)
						Sample size	199 (105) 141 [132-257] (58-450)	218 (98) 163 [152-299] (61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	114/ 990 (12%)	
						Max recruitment period	97/ 990 (9.8%)	
						Reached sample size cap	40/ 990 (4.0%)	
						Superiority at interim analysis	739/ 990 (75%)	
						Sample size	200 (107) 140 [132-252] (58-450)	
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	113/ 989 (11%)	
						Max recruitment period	92/ 989 (9.3%)	
						Reached sample size cap	38/ 989 (3.8%)	
						Superiority at interim analysis	746/ 989 (75%)	
						Sample size	200 (106) 140 [132-249] (58-450)	

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 994[1]	SRS 2 N = 972[1]
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	110/ 986 (11%)	69/ 967 (7.1%)
						Max recruitment period	74/ 986 (7.5%)	
						Reached sample size cap	47/ 986 (4.8%)	112/ 967 (12%)
						Superiority at interim analysis	755/ 986 (77%)	786/ 967 (81%)
						Sample size	200 (105) 146 [133-260] (58-450)	223 (98) 172 [154-309] (61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis		71/ 971 (7.3%)
						Reached sample size cap		128/ 971 (13%)
						Superiority at interim analysis		772/ 971 (80%)
						Sample size		218 (100) 162 [152-312] (61-400)
Treatment2								
	POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	589/ 994 (59%)	569/ 972 (59%)
						Max recruitment period	180/ 994 (18%)	
						Reached sample size cap	79/ 994 (7.9%)	282/ 972 (29%)
						Superiority at interim analysis	146/ 994 (15%)	121/ 972 (12%)
						Sample size	219 (132) 181 [79-338](59-450)	229 (136) 196 [73-400](61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 994[1]	SRS 2 N = 972[1]
						Futility at interim analysis	793/ 989 (80%)	754/ 974 (77%)
						Max recruitment period	97/ 989 (9.8%)	
						Reached sample size cap	43/ 989 (4.3%)	188/ 974 (19%)
						Superiority at interim analysis	56/ 989 (5.7%)	32/ 974 (3.3%)
						Sample size	181 (130) 134 [69-251](60-450)	195 (131) 153 [70-321](61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	214/ 990 (22%)	239/ 969 (25%)
						Max recruitment period	203/ 990 (21%)	
						Reached sample size cap	86/ 990 (8.7%)	316/ 969 (33%)
						Superiority at interim analysis	487/ 990 (49%)	414/ 969 (43%)
						Sample size	240 (133) 200 [133-360](60-450)	248 (128) 239 [151-400](62-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	132/ 989 (13%)	135/ 968 (14%)
						Max recruitment period	128/ 989 (13%)	
						Reached sample size cap	55/ 989 (5.6%)	241/ 968 (25%)
						Superiority at interim analysis	674/ 989 (68%)	592/ 968 (61%)
						Sample size	215 (118) 144 [132-317](60-450)	241 (117) 170 [152-395](63-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	102/ 990 (10%)	

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 994[1]	SRS 2 N = 972[1]
						Max recruitment period	85/ 990 (8.6%)	
						Reached sample size cap	33/ 990 (3.3%)	
						Superiority at interim analysis	770/ 990 (78%)	
						Sample size	197 (104) 140 [132-250](60-450)	
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	90/ 989 (9.1%)	
						Max recruitment period	65/ 989 (6.6%)	
						Reached sample size cap	30/ 989 (3.0%)	
						Superiority at interim analysis	804/ 989 (81%)	
						Sample size	191 (98) 139 [132-239](62-450)	
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis	580/ 986 (59%)	544/ 967 (56%)
						Max recruitment period	176/ 986 (18%)	
						Reached sample size cap	85/ 986 (8.6%)	308/ 967 (32%)
						Superiority at interim analysis	145/ 986 (15%)	115/ 967 (12%)
						Sample size	228 (142) 188 [80-362](59-450)	228 (138) 164 [73-400](61-400)
	POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Stopping reason		
						Futility at interim analysis		78/ 971 (8.0%)
						Reached sample size cap		146/ 971 (15%)

Group	POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS1 vs SRS2	Characteristic	SRS 1 N = 994[1]	SRS 2 N = 972[1]
						Superiority at interim analysis		747/ 971 (77%)
						Sample size		221 (103) 163 [152-314](63-400)

[1]n/ N (%); Mean (SD) Median [Q1-Q3] (Min-Max)

TABLE 8: MAIN RESULTS: SAMPLE SIZE (ASSUMING BINDING FUTILITY)

Group	Sample size cap	TRUE POR (POR 1,POR 2)	Efficacy stopping rule (POR eff - peff)	Futility stopping rule (POR fut - pfut)	Triggers not implemented at first interim analysis	SRS1 vs SRS2	total N = 999[1]	Treatment 1 N = 999[1]	Treatment 2 N = 999[1]
SRS 1									
	450	1.37,0.9	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	679 (265) 643 [430-916] [190-1,147]	209 (108) 169 [134-275] [58-450]	181 (130) 134 [69-251] [60-450]
	450	1.37,1.2	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	746 (274) 819 [440-1,034] [190-1,117]	196 (101) 141 [132-251] [58-450]	240 (133) 200 [133-360] [60-450]
	450	1.37,1.3	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	702 (266) 645 [433-1,004] [190-1,135]	199 (105) 141 [132-257] [58-450]	215 (118) 144 [132-317] [60-450]
	450	1.37,1.37	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	672 (256) 632 [430-889] [190-1,125]	200 (107) 140 [132-252] [58-450]	197 (104) 140 [132-250] [60-450]
	450	1.37,1.4	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	659 (250) 627 [430-870] [201-1,125]	200 (106) 140 [132-249] [58-450]	191 (98) 139 [132-239] [62-450]
	450	1.37,1	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	744 (275) 830 [439-1,030] [190-1,147]	200 (105) 146 [133-260] [58-450]	228 (142) 188 [80-362] [59-450]
	450	1,1	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	741 (310) 849 [431-1,036] [190-1,147]	211 (133) 170 [72-334] [58-450]	219 (132) 181 [79-338] [59-450]
SRS 2									
	400	1.42,0.9	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	738 (242) 807 [496-965] [201-1,269]	228 (100) 198 [155-319] [61-400]	195 (131) 153 [70-321] [61-400]
	400	1.42,1.2	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	794 (263) 826 [504-983] [201-1,281]	220 (99) 164 [153-297] [61-400]	248 (128) 239 [151-400] [62-400]
	400	1.42,1.3	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	772 (258) 812 [500-978] [201-1,281]	218 (98) 163 [152-299] [61-400]	241 (117) 170 [152-395] [63-400]
	400	1.42,1.42	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	733 (252) 788 [493-967] [201-1,281]	218 (100) 162 [152-312] [61-400]	221 (103) 163 [152-314] [63-400]
	400	1.42,1	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	779 (251) 828 [506-980] [201-1,271]	223 (98) 172 [154-309] [61-400]	228 (138) 164 [73-400] [61-400]
	400	1,1	1.1 - 0.83	1.1 - 0.8	FALSE	0.4 vs 0.6	796 (324) 876 [496-991] [197-1,279]	236 (138) 212 [73-400] [61-400]	229 (136) 196 [73-400] [61-400]

[1]Mean (SD) Median [Q1-Q3] [Min-Max]

9.1.2.3 Timing of the trial analyses

TABLE 9: MAIN RESULTS: TIME OF FIRST INTERIM ANALYSIS (MONTHS) AND TIME OF THE FINAL ANALYSIS (MONTHS) (ASSUMING BINDING FUTILITY)

POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS 1 SRS1 vs SRS2	Analysis	SRS 1 N = 994[1]	SRS 2 N = 972[1]
POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.74 (0.50) 15.72 [15.39-16.08] (14.13-17.66)
POR=1,1 (type I error)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	30 (10) 26 [20-42] (18-44)	26 (8) 24 [17-33] (15-40)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.73 (0.51) 15.72 [15.38-16.08] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=0.9(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	29 (7) 26 [25-32] (18-44)	25.1 (4.5) 23.1 [22.5-28.6] (15.6-39.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.73 (0.51) 15.72 [15.38-16.08] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.2(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	30 (7) 26 [25-32] (18-44)	25.6 (5.1) 23.1 [22.5-28.8] (15.6-40.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.44 [18.06-18.76] (16.55-20.05)	15.73 (0.50) 15.72 [15.38-16.08] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.3(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	30 (7) 26 [25-32] (18-44)	25.6 (5.1) 23.1 [22.5-28.8] (15.6-40.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.45 [18.06-18.76] (16.55-20.05)	
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.37(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	30 (7) 26 [25-32] (18-44)	

POR	Efficacy stopping rule	Futility stopping rule	Triggers not implemented at first interim analysis	SRS 1 SRS1 vs SRS2	Analysis	SRS 1 N = 994[1]	SRS 2 N = 972[1]
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.45 [18.06-18.76] (16.55-20.05)	
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.4(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	29 (7) 26 [25-32] (18-44)	
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)	18.42 (0.53) 18.45 [18.06-18.76] (16.55-20.05)	15.74 (0.50) 15.72 [15.39-16.09] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)	29 (7) 26 [25-32] (18-44)	25.2 (4.6) 23.1 [22.5-28.6] (15.6-39.0)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time first interim analysis (months)		15.73 (0.51) 15.72 [15.38-16.08] (14.13-17.66)
POR 1=1.37 (SRS1) POR 1=1.42 (SRS2) POR 2=1.42(power)	POR_eff=1.1 - peff=0.83	PORfut=1.1 - Pfut= 0.8	FALSE	0.4 vs 0.6	Time final analysis for treatment 1 (months)		25.6 (5.1) 23.1 [22.5-28.8] (15.6-40.0)

[1]Mean (SD) Median [Q1-Q3] (Min-Max)

The sensitivity analyses, where we vary the recruitment assumptions, OSFD distribution and SRS proportions are presented in the supplementary material ("Simulation report: Sensitivity analyses"). Briefly, it appears as though the (one-sided, pairwise) type I error remains controlled at approx. 10% and the (pairwise) power is >70% (for POR>=1.37 for SRS1 and POR=1.42 for SRS2) when these assumptions are varied.