MEDAL (MEDications in Acute Low back pain):

Design and Simulation Report

Generated by FACTS Core Engine for Time-to-Event Endpoint

Edited by Simon Gates

Version 1.1, March 31, 2022

Table of Contents

[2. Introduction 2](#_Toc99034116)

[2.1. Background 2](#_Toc99034117)

[2.2. Primary Endpoint 2](#_Toc99034118)

[2.3. Treatment Arms 2](#_Toc99034119)

[3. Statistical Modeling 2](#_Toc99034120)

[3.1. Primary Endpoint Model 3](#_Toc99034121)

[3.2. Quantities of Interest 3](#_Toc99034122)

[3.3. Posterior Probabilities 3](#_Toc99034123)

[3.4. Predictive Probability of Future Trial Success 4](#_Toc99034124)

[3.5. Target Doses 4](#_Toc99034125)

[3.6. Decision Quantities 5](#_Toc99034126)

[4. Study Design 5](#_Toc99034127)

[4.1. Timing of Interim Analyses 5](#_Toc99034128)

[4.2. Arm Dropping 5](#_Toc99034129)

[4.3. Criteria for Stopping Accrual 6](#_Toc99034130)

[4.3.1. Stopping for Expected Futility 6](#_Toc99034131)

[4.3.2. Stopping for Expected Success 6](#_Toc99034132)

[4.4. Final Evaluation Criteria 6](#_Toc99034133)

[5. Simulation Scenarios 6](#_Toc99034134)

[5.1. Virtual Subject Response Profiles 7](#_Toc99034135)

[5.1.1. Control Hazard Rates Profiles 7](#_Toc99034136)

[5.1.2. Hazard Ratio Profiles 7](#_Toc99034137)

[5.2. Accrual Profiles 7](#_Toc99034138)

[5.3. Dropout Profiles 8](#_Toc99034139)

[6. Operating Characteristics 9](#_Toc99034140)

[6.1. Overall 9](#_Toc99034141)

[6.2. Trial Outcomes 9](#_Toc99034142)

[7. Computational Details 10](#_Toc99034143)

# Introduction

## Background

This document describes the features of the simulated design for the MEDAL trial, including the statistical models, decision rules, and simulation scenarios as input into the (Fixed and Adaptive Clinical Trial Simulator) software. A set of basic operating characteristics for the simulations is also summarized; this will be expanded in future versions of this document.

The trial size and simulations described herein is for one of the three clinical diagnostic groups to be included in the trial, hence the total maximum size of the trial will be three times the maximum size used in this report. We have assumed that each group will be analysed in isolation, but in future work we will explore the use of hierarchical modelling across the three diagnostic groups, which may allow further improvement in trial performance.

## Primary Endpoint

The primary endpoint is time to recovery, which is the time to achieving a Patient Global Impression of Change (PGIC) score of ≥+4 points (at least ‘quite a lot better’) for 5 consecutive days i.e. a time-to-event outcome where an event is considered to be a positive outcome for a subject. All subjects will be monitored for events for 8 weeks.

## Treatment Arms

The trial will enroll up to a maximum of 1320 subjects, randomized among 6 treatment arms. Because of software limitations, the simulations also include a control arm; patients are randomized in a ratio 1 (control): 22 (each of six treatment arms), hence the total maximum number of recruits is 1330 (10 to the “control” arm and 220 to each treatment arm). The control arm plays no part in adaptive decisions or the conclusions of the trial.

The arms are labelled generically by their arm index as: 0 (Control), 1, 2, 3, 4, 5, 6.

**Table-1: Treatment Arms (automatically generated table not relevant to this trial)**

|  |  |
| --- | --- |
| Arm name | Arm strength |

# Statistical Modeling

This section describes the statistical modeling used in the design. The modeling is Bayesian in nature.

## Primary Endpoint Model

Let be the event time (in weeks) for the primary endpoint for the subject, where an event represents a positive outcome for the subject. We model the event times for the control arm as exponential:

where is the baseline hazard rate (events per week). The hazard rate is assumed to be constant over time and is modeled with prior:

The primary endpoint event times for the arms are also modeled as exponential, with hazard rates:

where is the log hazard ratio for arm relative to control. By convention, hazard ratios greater than one indicate treatment benefit.

The arms are modeled independently as:

Thus, for each arm is estimated separately using only data from that arm.

## Quantities of Interest

We define a number of quantities that will be tracked and may be used to make decisions during the trial.

## Posterior Probabilities

For each arm, we calculate the following quantities from the posterior:

* the probability that the primary endpoint hazard ratio on arm is less than arm\_1 by at least :
* the probability that the primary endpoint hazard ratio on arm is less than arm\_2 by at least :
* the probability that the primary endpoint hazard ratio on arm is less than arm\_3 by at least :
* the probability that the primary endpoint hazard ratio on arm is less than arm\_4 by at least :
* the probability that the primary endpoint hazard ratio on arm is less than arm\_5 by at least :
* the probability that the primary endpoint hazard ratio on arm is less than arm\_6 by at least :

## Maximum effectiveness targets

We consider the following targets:

* The “maximum effective dose” () is the arm with the greatest treatment effect (HR difference from 1). For each dose, we calculate the probability of being the :

## Decision Quantities

The above quantities are computed for each treatment arm (thus making them vector quantities). To facilitate decisions in the trial, we attach a particular treatment arm to the quantity that will be used for the decision (thus reducing the vector quantity to a scalar quantity). Throughout the trial, decisions may be based on the following quantities:

* Probability that current arm is worse than arm *d* by at least 0.4 (i.e. hazard ratio of 1.4 or more in favour of arm *d*), where *d* = 1, 2, 3, 4, 5, 6.
* Maximum (Pr(Max)) across all arms

# Study Design

## Timing of Interim Analyses

Interims will occur based on the number of subjects who have enrolled into the trial. Timings of interim analyses will vary if arms are dropped e.g. the second interim analysis will take place at approx. 147 recruits per arm, which will occur before 880 recruits (total) if arms are dropped following the first interim analysis.

The schedule for the 2 interims is specified in Table-2.

**Table-2: Interim Analysis Schedule**

|  |  |
| --- | --- |
| Interim | Subjects Enrolled |
| 1 | 440 |
| 2 | 880 |

## Arm Dropping

Randomization will occur in blocks of size 133, with subjects allocated in a ratio of 1:22:22:22:22:22:22 among the arms. The control arm plays no part in adaptive decision making and is included because of a software constraint which requires specification of a control groups.

Arm dropping criteria will be evaluated at interims 1, 2 . An arm will be a candidate for dropping if it meets any of the following criteria:

* Pr(*{d}-*{1}<-0.4) > 0.75
* Pr(*{d}-*{2}<-0.4) > 0.75
* Pr(*{d}-*{3}<-0.4) > 0.75
* Pr(*{d}-*{4}<-0.4) > 0.75
* Pr(*{d}-*{5}<-0.4) > 0.75
* Pr(*{d}-*{6}<-0.4) > 0.75

During the study, a maximum of 5 arms may be dropped, and no additional subjects will be allocated to those arms. The control arm may not be dropped at any time.  
More than one arm may be dropped at an interim, as long as the maximum number to drop has not yet been reached.

After arms are dropped, if any, the study is reduced in size i.e. the maximum size for one arm

## Criteria for Stopping Accrual

### Stopping for Futility

No early stopping criteria for futility have been defined for this trial.

### Stopping for Success

If at the second interim analysis, the best performing arm has a probability of being the best arm (pr(max)) of greater than 0.995, the trial can be stopped for that diagnostic group because there is a clear single best intervention.

If a success stopping rule is met at an interim analysis, then subject follow up will be discontinued, and the final evaluation criteria will be applied to the currently available data.

## Final Evaluation Criteria

No final success criteria have been defined for this trial.

No final futility criteria have been defined for this trial.

This is a Phase III trial, where the goal is to estimate the treatment effects and the comparative effectiveness of the different intervention, hence thresholds for success and futility are not appropriate.

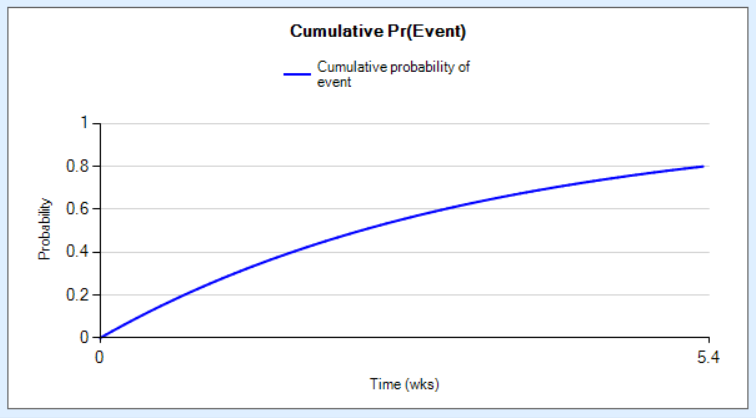
# Simulation Scenarios

We evaluate the proposed design through trial simulation. We hypothesize several possible underlying truths for the response in each arm, and can also explore the effects of trial execution variables such as accrual and dropout rates. For each of these scenarios, we generate data according to those truths and run through the design as specified above. We repeat this process to create multiple ``virtual trials" and we track the behavior of each trial. In this section, we describe the the parameters used to generate the virtual subject-level data.

## Virtual Subject Response Profiles

### Hazard Rates Profiles

We assumed a baseline hazard rate (for the non-used “control” group and non-effective intervention arms) of 0.3 events per week, which equates to approximately 5% of patients remaining non-recovered after 8 weeks and a median time to recovery of about 2 weeks.



### Hazard Ratio Profiles

We consider 4 profiles for which subject outcomes for the final endpoint are simulated to have hazard ratios as shown in Table-6.

**Table-6: Virtual subject response hazard ratios**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Scenario name** |  |  |  |  |  |  |
| **A** | 1 | 1 | 1 | 1 | 1 | 1 |
| **B** | 1 | 1.4 | 1 | 1.4 | 1 | 1 |
| **C** | 1 | 1 | 1.4 | 1 | 1 | 1 |
| **D** | 1 | 1.2 | 1 | 1.4 | 1 | 1 |
| **E** | 1 | 1.2 | 1 | 1.2 | 1 | 1 |
| **F** | 1 | 1 | 1 | 1.2 | 1 | 1 |

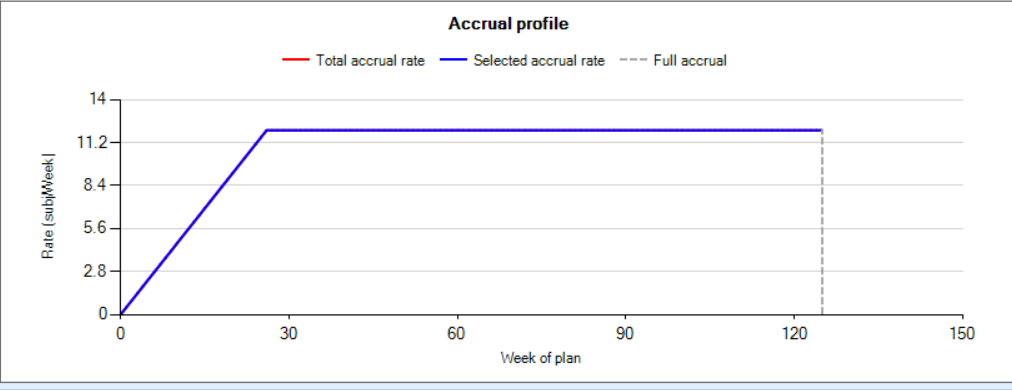
## Accrual Profiles

We simulate the random arrival of subjects into the trial from a Poisson process with the mean weekly rates specified in Table-7. Within each accrual profile, there may be differential recruitment rates over time and across regions. Thus for each region, we specify:

* the mean number of subjects per week at peak accrual,

**Table-7: Accrual Profiles**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Profile Name | Region Index | Peak Rate | Start Week | Ramp Up | Ramp up Complete | Ramp Down | Start Ramp Down | Ramp Down Complete |
| Acc 1 | 1 | 12 | 0 | Yes | 26 | NA | NA | NA |



## Dropout Profiles

We consider the following dropout profile:

*Dropout 1*

We simulate subjects dropping out of the trial with the dropout rates (per week) shown in Table-8

**Table-8 : Dropout Profiles for Dropout 1**

|  |  |
| --- | --- |
| Arm | 0 - weeks |
| arm\_1 | 0.01 |
| arm\_2 | 0.01 |
| arm\_3 | 0.01 |
| arm\_4 | 0.01 |
| arm\_5 | 0.01 |
| arm\_6 | 0.01 |

# Operating Characteristics

For the scenarios described above, we simulate multiple virtual trials and track the behavior of each trial, including the final outcome of the trial, the estimated mean response, etc. The results in this section are summarized across all simulated trials for each scenario.

NB the simulations reported in this report may differ slightly from those included in the NIHR HTA application, as they derive from a different run of the simulation program. The overall conclusions are the same.

## Overall

This section gives a high-level description of the operating characteristics. Table-9 shows the following information per scenario:

* N sim: the number of simulated trials
* E[N]: the expected sample size
* Pr(stop early for success): the proportion of trials that met the final success criteria. For these simulations, this was the proportion that were stopped early for meeting the success criterion at the second interim analysis.
* E[duration]: the expected duration of the trial in weeks.

**Table-9: Overall Operating Characteristics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hazard Ratio Scenario** | **N sim** | **E[N]** | **Pr(Stop early for success)** | **E[duration]** |
| A | 1000 | 1321.7 | 0.000 | 131.5 |
| B | 1000 | 1131.7 | 0.009 | 115.4 |
| C | 1000 | 1136.3 | 0.150 | 114.8 |
| D | 1000 | 1173.1 | 0.052 | 118.6 |
| E | 1000 | 1283.4 | 0.001 | 128.3 |
| F | 1000 | 1290.8 | 0.007 | 128.9 |

Table 10 (below) shows the proportion of each arm stopped early for success or dropped at either of the interim analyses. In scenario A, where no interventions are better than any others, about 1% of each arm was incorrectly stopped early for inferiority. Stopping for success did not occur at all in this scenario. Where there was one effective arm (C and F), that arm was very rarely (< 1%) dropped for ineffectiveness, even with a hazard ratio as low as 1.2 (scenario F). With a true hazard ratio of 1.4 (C), 14% of simulations were able to stop early for “success,” identifying the single best intervention with very high probability. This scenario also allowed ineffective arms to be dropped early with probability 21% - 24%, leading to a smaller overall sample size.

**Table 10: Proportion of each arm stopped for success or dropped (at either interim analysis). Red indicates arms with hazard ratio of 1.2 or 1.4.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Scenario | Mean sample size | Proportion stopped early for success | Proportion of each arm dropped at 1st or 2nd interim analysis | | | | | |
|  |  |  | Arm 1 | Arm 2 | Arm 3 | Arm 4 | Arm 5 | Arm 6 |
| A | 1311 | 0 | 0.012 | 0.011 | 0.006 | 0.009 | 0.010 | 0.010 |
| B | 1112 | 0.010 | 0.394 | 0.006 | 0.396 | 0.007 | 0.378 | 0.374 |
| C | 1106 | 0.141 | 0.216 | 0.241 | 0 | 0.218 | 0.234 | 0.217 |
| D | 1155 | 0.045 | 0.276 | 0.047 | 0.277 | 0.001 | 0.253 | 0.250 |
| E | 1260 | 0.002 | 0.085 | 0.003 | 0.090 | 0.007 | 0.085 | 0.091 |
| F | 1273 | 0.005 | 0.054 | 0.053 | 0.054 | 0.002 | 0.057 | 0.058 |

Where there were two equally good interventions (B and E), the best interventions were very rarely stopped early, but other arms were, with probabilities u to 40% in arm B and 10% in arm E. Stopping for success was rare in these scenarios, which is expected because if two interventions are equally good it is hard to distinguish between them so it is appropriate to continue randomizing between them up to the maximum sample size for each arm (220).

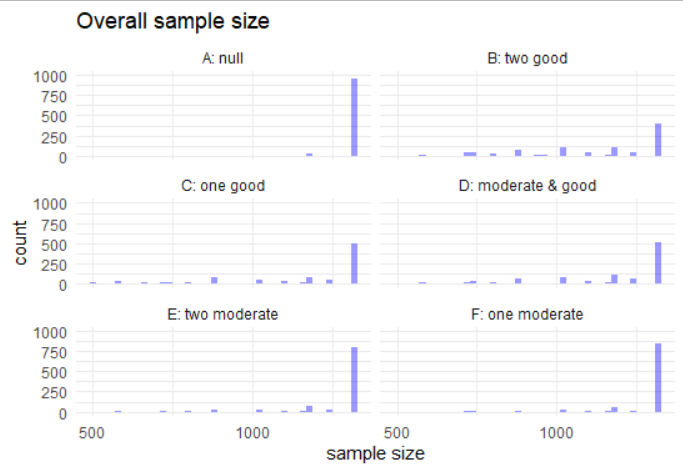
Table 11 (below) summarises the modeled probability of each arm being thebest at the end of the trial.

**Table 11: Proportion of simulations in which each arm was the best (i.e. highest prob(max)) at the end of the trial. Red indicates arms with hazard ratio of 1.2 or 1.4.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Scenario | Proportion of simulations in which each arm had the highest probability of being the best | | | | | |
|  | Arm 1 | Arm 2 | Arm 3 | Arm 4 | Arm 5 | Arm 6 |
| A | 0.164 | 0.170 | 0.168 | 0.185 | 0.164 | 0.149 |
| B | 0 | 0.465 | 0 | 0.535 | 0 | 0 |
| C | 0 | 0 | 0.997 | 0 | 0.002 | 0.001 |
| D | 0 | 0.053 | 0 | 0.947 | 0 | 0 |
| E | 0.004 | 0.468 | 0.004 | 0.513 | 0.004 | 0.007 |
| F | 0.024 | 0.021 | 0.016 | 0.901 | 0.019 | 0.019 |

Figure 1 (below) summarises the overall sample size of the trial in 1000 simulations. When all interventions are the same (A), almost all simulations continue to the maximum sample size, whereas when there are one or two better interventions with a hazard ratio of 1.4 (B and C), less than half reach the maximum sample size, with the remainder stopping at sample sizes between 500 and 1320.

**Figure 1: histogram of overall sample sizes in 1000 simulations**



# Computational Details

This report reflects the design parameters contained within the **MEDAL\_TTE\_1320.facts** file. The simulations were run using FACTS (Berry Consultants, LLC, Austin, TX) version 6.4.1. Table-12 shows the computational details for each scenario, including the starting date and time, the length of the MCMC chain, the random number seed, and the trial at which the simulation started. The R software package was used to summarize the simulation output and to create tables for this report.

**Table-12: Computational Details**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Accrual | Dropout | Hazard Ratio | Control Hazard Rates | Date Time | MCMC Burn-in | MCMC Length | Random Seed | Starting Trial |
| Acc 1 | Dropout 1 | 1 | good | 03/24/2022 10:40:44 | 1000 | 2500 | 417017577 | 1 |
| Acc 1 | Dropout 1 | 1 | mod | 03/24/2022 10:40:44 | 1000 | 2500 | 417017577 | 1 |
| Acc 1 | Dropout 1 | 2 | good | 03/24/2022 10:40:43 | 1000 | 2500 | 417017577 | 1 |
| Acc 1 | Dropout 1 | 2 | mod | 03/24/2022 10:40:44 | 1000 | 2500 | 417017577 | 1 |
| Acc 1 | Dropout 1 | mod&good | CHaz 1 | 03/24/2022 10:40:44 | 1000 | 2500 | 417017577 | 1 |
| Acc 1 | Dropout 1 | null | CHaz 1 | 03/24/2022 10:40:43 | 1000 | 2500 | 417017577 | 1 |