Statistical Analysis Appendix and Adaptive Design Report for SPRY-Metformin Domain

1.0 Introduction

SPRY-Metformin is a randomized control trial comparing the effectiveness of different doses and durations of metformin to placebo for nondiabetic patients with elective surgeries. In particular, we will evaluate 3 doses of metformin (500, 1000 and 1500mg) as well as 3 levels of pre-op duration of metformin (short, 7-28 days; intermediate, 29-90 days; and long, 90 days). Patients will be randomized to one of the three metformin doses or placebo but will not be randomized to the pre-op duration. Pre-op duration will be observed based on the timing of the first pre-op visit.

The primary endpoint to determine efficacy of metformin relative to placebo is hospital free days (HFD) at day 90 after the surgical encounter after administration of metformin vs. placebo. HFD at day 90 is an ordered categorical variable that takes on discrete integer values from -1 to 90 and is calculated as 90 minus the number of days of the index stay and the number of days readmitted within the 90-day time period following the surgical encounter. If mortality occurs within the 90-day time period, the patient is given an HFD value of -1 (ordered to be a worse outcome than being in the hospital for all 90 days).

There will be a maximum of 2000-2500 patients randomized in the trial. Within each of the 3 pre-op durations, patients will initially be randomized $\sqrt{3}$:1:1:1 to placebo and the 3 doses of metformin until a total of 500 patients have been randomized across all pre-op durations and followed for 90 days. Afterwards, interim analyses will occur sequentially after an additional 500 patients have been followed for 90 days. At each interim analysis, the trial can be stopped early for demonstrating efficacy of one of the metformin doses compared to placebo (see Section 3.1). If the trial has not stopped for success and continues enrolling, within each pre-op duration doses can be dropped for futility and responsive adaptive randomization will be used to randomize patients preferentially to the best performing metformin doses of all of the remaining doses within that pre-op durations (see Sections 3.2-3.3). The trial can stop enrolling patients within a pre-op duration if all metformin doses have been stopped within that duration for futility (see Section 3.2). Finally, at the interim when 2000 patients have been randomized across all pre-op durations and followed for 90 days, the maximum sample size could be increased from 2000 to 2500 (see Section 3.4).

2.0 Statistical Modeling

Inferences and quantities of interest used for response adaptive randomization, success or futility of metformin doses, and increasing the maximum sample size in this trial are based a Bayesian ordinal logistic regression model. The ordinal logistic regression model

Reitz KM, *et al. BMJ Open* 2020; 10:e037690. doi: 10.1136/bmjopen-2020-037690

accounts for underlying differences in the expected 90-day HFD distribution depending on surgical procedure or strata of the patient but assumes a common odds ratio treatment effect across the surgical strata. The odds ratio shift within an ordinal logistic regression model can be thought of similarly to an odds ratio within a logistic regression analysis of a dichotomous endpoint. Within ordinal logistic regression, we are simply performing multiple logistic regression analyses (one for each possible dichotomization of the data) and providing a weighted average of the odds ratios across these different dichotomizations. The assumption of a common odds ratio treatment effect across the different surgical subtypes translates into different absolute differences in the mean hospital free days within each surgical strata. For a common odds ratio across surgical strata, the larger the expected HFD within the strata the smaller the absolute mean difference in HFD between treatment and control.

In this setting, the Bayesian analysis makes use of non-informative prior distributions with regards to HFD distributions for each surgical strata and in this regard is very similar in nature to a frequentist ordinal logistic analysis. However, we chose to use a Bayesian analysis over a frequentist approach to allow for borrowing of information on the treatment effect across different doses and durations. This borrowing is done in the Bayesian setting by placing a hierarchical prior distribution on the treatment effects across all doses, all durations and the interactions between them.

2.1 Bayesian Ordinal Logistic Regression

Throughout we assume for patient *i,* Y_i is the observed 90-day HFD, $g(i)$, is the surgical strata from 1:G, *d(i)* is the pre-op duration from 1:3 with 1 = short, 2 = intermediate, and 3=long, and *t(i)* is the intervention from 1:4 with 1 = placebo, 2 = 500mg, 3 = 1000mg, and 4 = 1500mg.

A Bayesian ordinal logistic regression model is used to estimate the effect of dose and duration of metformin on the distribution of HFD under placebo adjusting for expected differences given the surgical type/strata. The ordinal scale parameterization is a generalized version of the dichotomous parameterization where we model all cumulative probabilities of 90-day HFD being less than or equal to a cut point *c*, where c=–1,…, 89. Given each cut point *c,* we denote the 91 dichotomized versions of 90-day HFD for patient *i* as $Y_{c,i}$ where $Y_{c,i} = 1$ if 90-day HFD is in *[-1, c]* and $Y_{c,i} = 0$ if 90-day HFD is in [c+1,90] for *c=-1,…, 89.* $Y_{c,i}$ is then modeled throughout as:

$$Y_{c,i} \sim Bernoulli(\phi_{c,i}), c = 1 \dots 89;$$
$$logit(\phi_{c,i}) = \gamma_c + \mu_i;$$

where μ_i is a patient-specific mean function and γ_c is common across all patients.

The subject-specific mean function is as follows:

$$\mu_i = \alpha_{g(i)} + \theta_{t(i),d(i)}, i = 1 \dots N.$$

Reitz KM, *et al. BMJ Open* 2020; 10:e037690. doi: 10.1136/bmjopen-2020-037690

Within this model we assume that the underlying distribution of HFD is different within each stratum, *g*, and these differences across strata can be explained by a proportional log-odds ratio shift in the HFD distribution, α_g. Furthermore, we assume that the effects of each intervention within each pre-op duration are constant across strata and can be explained by a proportional log-odds ratio shift in the HFD distribution $\theta_{t,d}$. Where a log-odds ratio $\theta_{t,d} < 0$ results in an increase in expected HFD. For identifiability we assume the effect of placebo across all durations is zero, $\theta_{1,d} = 0 \; for \; all \; d = 1{:}3$. As such, the values of the inverse logit of γ_c define the cumulative probabilities for each HFD value under placebo, common across pre-op durations, and averaged across all strata. For all doses of metformin, we assume that the log-odds ratio of the effect of the dose is dependent on the pre-op duration and takes on the following form:

$$\theta_{t,d} = \beta_t + \kappa_d + \delta_{t,d} \;\; for \; t > 1, t = 1 \ldots 4, d = 1 \ldots 3.$$

Here, β_t is the log-odds ratio due to the dose, κ_d is the log-odds ratio due to the duration and $\delta_{t,d}$ is an interaction between dose and duration.

2.2 Model Priors

The prior distribution of γ_c is specified on the probability scale:

$$\pi \sim Dirichlet(\alpha_{-1}, \cdots \alpha_{90});$$
$$\gamma_c = logit\left(\sum_{i=-1}^{c} \pi_i\right), c = 1 \ldots 89;$$

with hyper-parameters, α_h, specified based on the observed rates of HFD across all strata in pre-trial data (discussed in Section 4) and providing 1 patient worth of information so that $\sum_{h=-1}^{90} \alpha_h = 1$.

For the strata-specific log-odds ratios we place a normal prior distribution with mean 0 and standard deviation 2:

$$\alpha_g \sim N(0, 2^2), g = 1 \ldots G.$$

Within pre-trial data (discussed in Section 4), the standard deviation of the log-odds ratios across surgical types/strata was estimated to be 1.5.

We assume a hierarchical distributions for the dose-effects and duration-effects each centered around a common mean so there is borrowing of information across doses and durations:

$$\beta_t \sim N\left(\mu_\beta, .5^2\right); \; \mu_\beta \sim N(0,1), t = 1 \ldots 4;$$
$$\kappa_d \sim N(\mu_\kappa, .5^2); \; \mu_\kappa \sim N(0,1), d = 1 \ldots 3.$$

Reitz KM, *et al. BMJ Open* 2020; 10:e037690. doi: 10.1136/bmjopen-2020-037690

Finally, we assume that the interaction between dose and duration has a normal prior distribution with mean 0 and standard deviation .2 to limit the amount of deviation of the overall effect, $\theta_{t,d}$, from the two additive effects.

2.3 Quantities of Interest

The following statistical quantities are used in the design of the trial and will be summarized at the conclusion of the trial. The posterior distribution of all model parameters is calculated using MCMC. The algorithm allows the generating of M (ex. 100,000) draws from the joint posterior distribution for all model parameters.

2.3.1 Summaries of Treatment Effect

The effect of each dose, *t*, and duration, *d*, will be summarized by reporting the posterior mean and 95% CI of the odds ratio, $\exp(\theta_{t,d})$, (common across all surgical strata). Additionally, we will translate the posterior mean odds ratio into expected mean differences in HFD for each surgical subtype enrolled in the trial. Finally, we will report raw mean (and SD) differences in HFD for each surgical subtype, under each dose and duration. These raw estimates will not take into account the borrowing of information across doses and durations.

2.3.2 Probability beat placebo by CSD

To determine if a dose should be dropped within a duration or if we should increase the sample size at N=2000, we summarize the posteriority probability that each dose, *t*, and duration, *d*, of metformin is superior to placebo by some clinically significant difference (CSD). The CSD is defined as an odds ratio of .8. Thus, we are interested in the probability $\exp(\theta_{t,d}) < .8$. This quantity is calculated from the M samples of the posterior distribution of the effect of each dose and duration, $\theta_{t,d}$, by reporting the proportion of posterior samples in which the odds ratio, $\exp(\theta_{t,d})$ is less than .8:

$$\Pr\left(\exp(\theta_{t,d}\right) < .8 \mid \mathrm{Y}) = \frac{1}{M}\sum_{m=1}^{M}(\exp(\theta_{t,d}) < .8), t = 1 \ldots 4, d = 1 \ldots 3.$$

2.3.4 Probability of Optimal Dose within each Duration

Within a pre-op duration, we will use response adaptive randomization to allocate the next set of patients to all doses that have not been stopped for futility based the posterior probability that each dose is optimal within each pre-op duration. This quantity is calculated from the M samples of the posterior distribution of the effect of each dose within each duration, $\theta_{t,d}$, by reporting the proportion of posterior samples in which the log odds ratio for dose *t*, $\theta_{t,d}$ is the min observed effect across all three metformin doses *t=2:4* with duration *d*:

Reitz KM, *et al. BMJ Open* 2020; 10:e037690. doi: 10.1136/bmjopen-2020-037690

$$O(t,d) = \frac{1}{M}\sum_{m=1}^{M} I[\theta_{t,d} < \theta_{j,d} \; for \; all \; j \neq t], t = 1 \ldots 4, d = 1 \ldots 3.$$

2.3.5 Probability of Superiority

To determine if the trial should stop early for success at any interim or if the trial is successful at the final analysis, we summarize the posteriority probability that each dose of metformin is superior to placebo. For the superiority analysis, we estimate the effect of each dose of metformin by pooling across all actively enrolling durations. This is achieved by using the model described in Section 2.1 with the additional assumption that $\theta_t = \theta_{t,1} = \theta_{t,2} = \theta_{t,3}$. The posterior distribution of the pooled effect of each dose, θ_t, is this estimated by calculating M samples of the posterior distribution using only data from the actively enrolling doses within each duration. The probability of superiority of each dose relative to placebo is then calculated as the proportion of the M samples with θ_t less than zero:

$$\Pr(\theta_t < 0 \mid \mathrm{Y}) = \frac{1}{M}\sum_{m=1}^{M} \theta_t < 0 \, , t = 1 \ldots 4.$$

2.4 Missing Data

All missing data will be imputed based upon the median observed 90-day HFD value for each treatment arm and preoperative duration. Sensitivity analyses will utilize the following different imputation strategies that do not assume missing at random (MAR):

- Impute all missing values as the median observed 90-day HFD value under the placebo group.
- Impute all missing values as the worse observed 90-day HFD value within that surgical strata.

2.5 Heterogeneity of Treatment Effects

The heterogeneity of treatment effects across different key patient subgroups will be explored by allowing the common odds ratio per dose and duration, exp $(\theta_{t,d})$, to be subgroup dependent. Subgroups of interest include: Surgical specialty (see Table 4.1), operative stress level as defined by the Operative Stress Score,[1] surgical subtype, age category, and frailty based upon the prospectively calculated Revised Analysis Index.

3.0 Interim Analyses and Trial Adaptations

Before interim analyses begin, patients will be randomized √3:1:1:1 to placebo and the three doses of metformin within each pre-op duration. Interim analyses will then begin when 500 total patients across all doses and durations are randomized and have been followed for 90 days and will continue after every additional 500 patients have been

followed for 90 days. Thus, there are 4 total interims at 500, 1000, 1500, and 2000 patients with 90-day follow-up and a final analysis when 2500 patients have been followed for 90 days. At each interim we allow the following adaptations:

- Success
- Dose / Duration Dropping
- Response Adaptive Randomization

3.1 Success

Success will be declared at an early interim or at the final analysis, and the trial will stop if the posterior probability of superiority of any dose of metformin relative to placebo defined in Section 2.3.3 is greater than a pre-defined interim-specific threshold. The thresholds for each interim are reported in Table 3.1.1 and are based on an O'brien Fleming spending function assuming a maximum sample size of 2500:

Table 3.1.1: Success Thresholds

Analysis	**500**	**1000**	**1500**	**2000**	**2500**
Success Threshold	***.9999***	***.9999***	***.9985***	***.9950***	***.9894***

3.2 Dose / Duration Dropping

Metformin doses will be dropped within a duration based on the probability of futility defined in Section 2.3.1. Specifically, for dose *t* in duration *d* if

$$\Pr\left(\exp\left(\theta_{t,d}\right) < .8 \mid \mathrm{Y}\right) < .15, t = 1 \ldots 4, d = 1 \ldots 3;$$

dose *t* will be dropped in duration *d* and patients within that duration will no longer be randomized to that dose.

We require an additional order restriction on dose dropping so that a dose must be dropped first in the short duration, then the intermediate duration then the long. Therefore, a dose cannot be dropped in the intermediate duration until it has first been dropped in the short and cannot be dropped in the long duration until it has first been dropped in the short and intermediate.

Enrollment to a pre-op duration will be stopped if all doses within that duration have been stopped and the trial will stop for futility if all pre-op durations have been stopped.

3.3 Response Adaptive Randomization within Durations

Reitz KM, *et al. BMJ Open* 2020; 10:e037690. doi: 10.1136/bmjopen-2020-037690

Within each pre-op duration of metformin, we will use response adaptive randomization to allocate patients to the most optimal dose of metformin within that pre-op duration. Initial randomization is set to $\sqrt{3}$:1:1:1 to placebo and the 3 doses of metformin within each duration. This allocates approximately .366 percent of the patients to placebo. This percentage allocation to placebo will be maintained throughout the course of the trial. However, after the first interim analysis, the remaining .634 percent of patients will be allocated to metformin doses within each duration that have not been dropped for futility and preferentially based on the probability that the dose is optimal within the duration defined in Section 2.3.2 and renormalized over the currently enrolling doses.

3.4 Increasing maximum sample size to 2500

At the interim analysis when 2000 patients are randomized and followed for 90 days the maximum sample size will increase to 2500 if at least one dose within one pre-op duration meets the following criteria:

$$\Pr\left(\exp\left(\theta_{t,d}\right) < .8 \mid \mathrm{Y}\right) > .50, t = 1 \ldots 4, d = 1 \ldots 3.$$

After 2000 patients have been randomized and are waiting to be followed for 90 days, enrollment will continue until the interim analysis takes place. If the above criteria is met, enrollment will continue to a maximum of 2500. If the above criteria is not met, enrollment will stop.

4.0 Clinical Trial Simulations

Clinical trial simulations are used to provide example trial results, to optimize clinical trial design (best thresholds for early success, dose dropping, and futility stopping) and to determine the sample size needed within this trial to obtain at least 80% power for a clinically meaningful treatment effect and a one-sided 2.5% type I error under the null distributions. Simulations were provided under a wide range of clinical trial parameters to optimize the design with the design team. Operating characteristics are provided for the final design herein.

To create realistic clinical trial simulations, we obtained pre-trial data from patients within the UPMC electronic health records who had received an in-patient elective surgery and met the additional inclusion/exclusion criteria:

- Inclusion:
  - Age > 60 or RAI > 30 or CCI >2
  - Surgery performed in either PUH or SHY hospitals
- Exclusion:
  - Diabetes or previous metformin use
  - Had one of the following surgery types:
    - Minimally invasive cholecystectomy
    - Irrigation and debridement of a wound
    - Hyst. Total abdomen

- Vaginal Hyst.
- Sleeve Gast.

This resulted in data from 16,932 patients across 376 surgery types. Table 4.1 provides summaries of the data by clustering each surgery type into one of 14 surgical specialties. In particular, for each surgical specialty we report: total number of patients, total number of surgical types, mean and median HFD, and 90-day mortality rates.

Table 4.1: Summary Pre-Trial Data

	Total N	Surgical Procedures/ Strata	Mean HFD	Median HFD	Mort. Rate
Total	**16832**	**376**	**79.5**	**86.0**	**0.05**
ORTHO	3849	72	83.2	87.0	0.03
SPINE	2884	25	83.6	87.0	0.02
CARDIAC	1979	34	75.5	83.0	0.07
GENERAL	1692	52	70.9	82.0	0.10
UROLOGY	1221	21	85.2	88.0	0.01
THORACIC	1130	35	76.7	84.0	0.06
NEURO	1099	35	78.0	87.0	0.08
VASCULAR	1043	39	77.7	86.0	0.07
HPB	729	16	78.6	84.0	0.03
COLORECTAL	707	20	77.0	84.0	0.04
ENT	334	8	79.6	86.0	0.04
TRANSPLANT	136	8	71.0	81.5	0.01
GYNE	15	7	80.1	86.0	0.07
BARIATRIC	14	4	73.3	80.0	0.07

4.1 Virtual Patient Simulation

Within each simulation, we assumed that the SPRY trial would enroll subjects from all strata that had at least 50 subjects in the pre-trial data (77 total) with the proportion of patients within each enrolling stratum estimated from the pre-trial data. We also assume that the HFD distribution per strata under placebo is the same as what was observed in the pre-trial data. Finally, we assume treatment effects for each metformin dose can be summarized as a common percent reduction in the mean hospital days (HD) across all strata. This treatment effect is assumed to be 0% for all null scenarios and a maximum of 15% for all alternative scenarios.

The trial was powered assuming a common treatment effect of 15% reduction in mean hospital days across all surgical subtypes. The common treatment effect is specified as a percent reduction in mean hospital days to help elicit the minimal clinically meaningful treatment effect from the trial design team. The reduction of 15% in hospital days is thought to be the minimal clinically meaningful treatment effect within this patient population. This common percent reduction across surgical subtypes, results in different absolute effects in mean hospital days depending on the surgical subtype (see Table 4.2). In particular, for one of the most common surgical subtypes of Total Knee Arthroplasty, this would result in a half of a day reduction in hospital days (3.4 days in hospital vs. 2.9

days). In comparison, under Endovascular aortic repair, the expected reduction in hospital days is 1.6 (10.8 days in hospital vs. 9.2 days). A percent reduction that that is at least 15% would result in a savings ranging from 0.2 – 3.9 hospital days for the 10 most common surgical types.

To obtain a common percent reduction in mean HD across all strata within our simulations, we find the strata-specific odds ratio shift under treatment relative to the empirical HFD distribution under placebo that results in the assumed common percent reduction in HD per strata.

For example, Figure 4.1 plots the assumed HFD distribution under placebo and under a 15% reduction in HD for the most common surgical type, Total knee arthroplasty. Within the pre-trial data there were 1115 patients who received a total knee arthroplasty. The empirical HFD distribution observed in the pre-trial patients and assumed for placebo within this stratum is plotted in blue with approximately 10% of patients having 89 HFD, 35% with 88 HFD and 29% with 87 HFD. Across all patients, the mean HFD is 86.6. To achieve a treatment effect of a 15% reduction in HD (plotted in green) we would need an odds ratio shift in the treatment distribution relative to placebo of .62. This would result in a mean reduction in HD of .5. This would shift approximately 15% of patients under treatment to 89 HFD, 42% to 88 HFD and 25% to 87 HFD.

Similar summaries for the 10 most common surgical types are provided in Table 4.2.

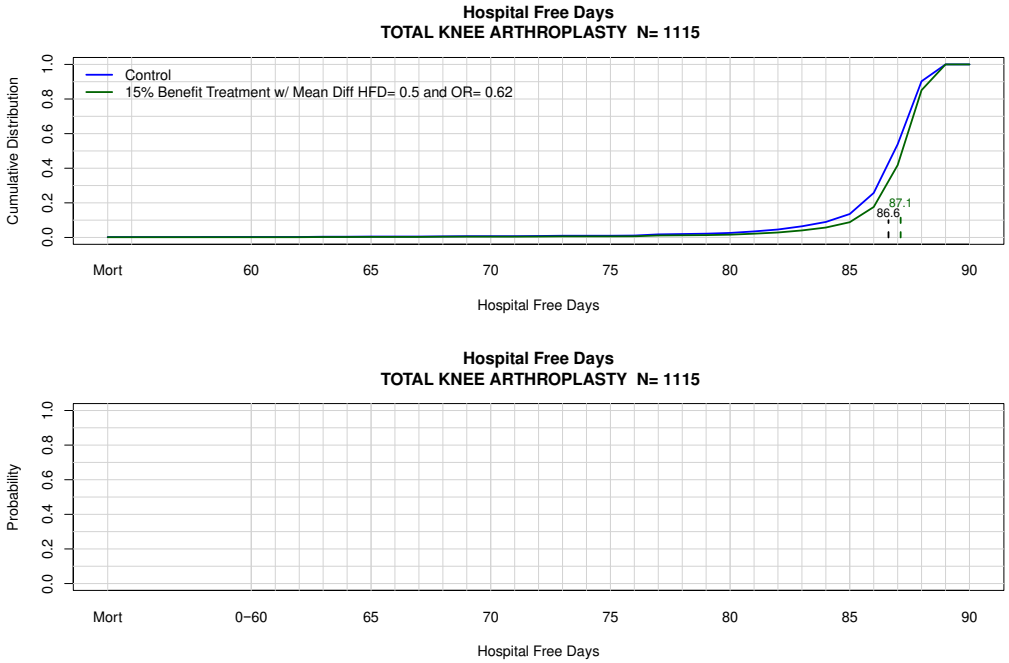

Figure 4.1: Example Strata-Specific HFD distribution under placebo vs. treated with a 15% reduction in HD for Total Knee Arthroplasty.

Table 4.2: Summary of Control Distributions and Treatment effects for 10 most common surgical types.				
	Prop. Overall	**Mean HFD Control**	**Mean Diff. Under Common 15% Reduction in HD**	**Odds-Ratio Shift Under Common 15% Reduction in HD**
Total Knee Arth.	0.08	86.6	0.5	0.62
Spine Post. Fuse Internal Fix.	0.07	82.8	1.1	0.71
Total Hip Arth.	0.05	85.6	0.7	0.72
Endo. Aortic Valve Replace	0.03	79.2	1.6	0.78
Spine Ant. Cervical Dissect. and Fuse	0.03	86.2	0.6	0.77
Spine Post. Lumbar or Thoracic	0.03	84.7	0.8	0.77
MIS Partial Pulmonary Lobectomy	0.02	83.2	1.0	0.73
Prostatectomy Lap. Robotic Assist.	0.02	88.6	0.2	0.50
Laparotomy	0.02	64.1	3.9	0.76
Total Hip MIS 2 Incisions	0.02	88.0	0.2	0.76

5.0 Example Trials

We provide example data and results for two simulated example trials. In particular, for each interim in each example trial we provide a plot of the data and results (ex. Figure 5.1.1). Each plot shows the following:

- Top Left: Allocation to each dose and the number of patients within each duration for each dose.
- Top Middle: Mean estimates (circles) and CI for the ORs for each dose and duration of metformin as well as pooled for each dose (above the P and in grey) across all actively enrolling durations. The confidence intervals show the lower .15 quantile so that if the lower bar goes above .8 the dose may stop for futility and the upper Xth quantile where X is interim specific success threshold based on the success rules provided in Table 3.1.1 so that if the upper bar goes below 1 for the pooled estimate, the dose will be declared a success. Raw OR values are provided plotted as stars.
- Top Right: The new allocation probabilities within each duration for placebo and the 3 metformin doses.
- Bottom: Cumulative probabilities of observing each HFD value or less for Placebo and each dose of metformin averaged across all durations and separately within each duration. As the curves move down and to the right, the expected HFD is increasing and the number of expected HD is decreasing.

5.1 Example Trial 1

Figure 5.1.1 shows results from the first interim analysis when 500 patients have 90-day data. Approximately 200 patients have been allocated to placebo and 100 to each of the 3 metformin doses. Estimates for the OR of all doses (500, 1000 and 1500) given at the short duration are 1.2 or greater, all have a posterior probability that the OR < .8 less than 15%, and all are stopped for futility. Thus, the trial stops enrolling in the short duration. All doses are still enrolling in the medium and long durations. Within the intermediate duration the 1500mg dose has an OR estimated around .75, and the 1000 and 500mg have an OR estimated around .85. Therefore, the new allocation probabilities are weighted towards the 1500mg dose within the intermediate duration. Within the long duration the 1500 and 1000mg doses have an OR estimated around .6 and are preferentially allocated to over the 500mg dose which has an estimated OR of .85.

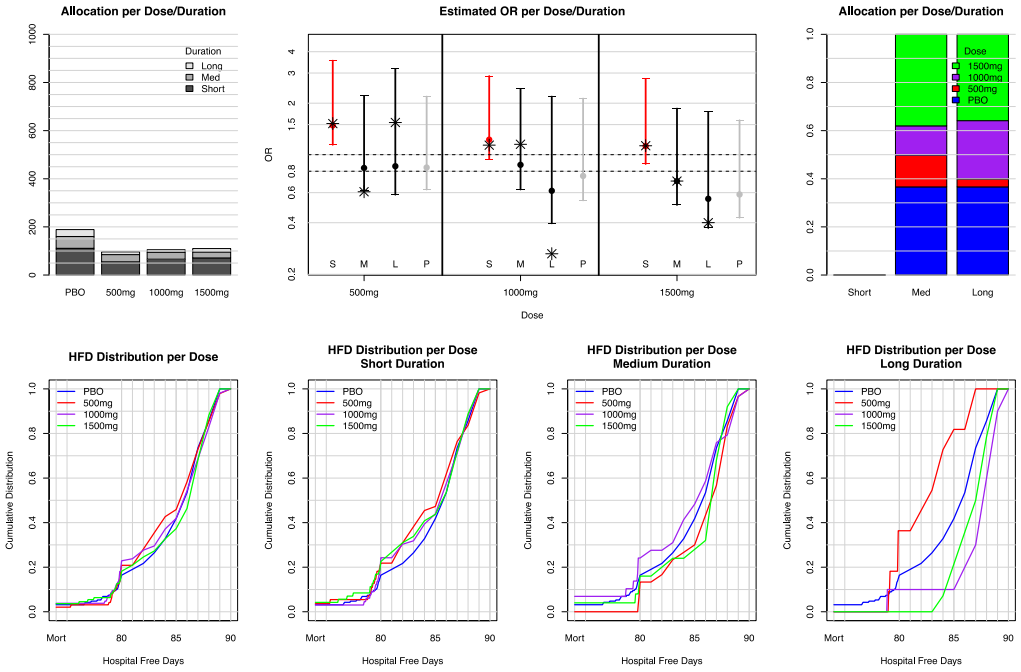

Figure 5.1.1: Example Trial 1; Interim N=500

Figure 5.1.2 shows results from the second interim analysis when 1000 patients have 90-day data. Approximately 375 patients have been allocated to placebo, 150 to 500mg, 200 to 1000mg and 300 to 1500mg. No new patients have been enrolled in the short duration. Within the intermediate duration the 1500mg dose has an OR estimated around .70, and the 1000 and 500mg have an OR estimated around .90. Therefore, the new allocation probabilities are weighted towards the 1500mg dose and away from the 1000 and 500mg dose within the intermediate duration. Within the long duration the 1500mg and 1000mg doses have an OR estimated around .65 and .75 respectively and are preferentially allocated to over the 500mg dose which has an estimated OR greater than 1. The 500mg dose in the long duration has less than a 15% posterior probability of having an OR < .8. However, it is not stopped since the intermediate duration has not stopped yet for this dose.

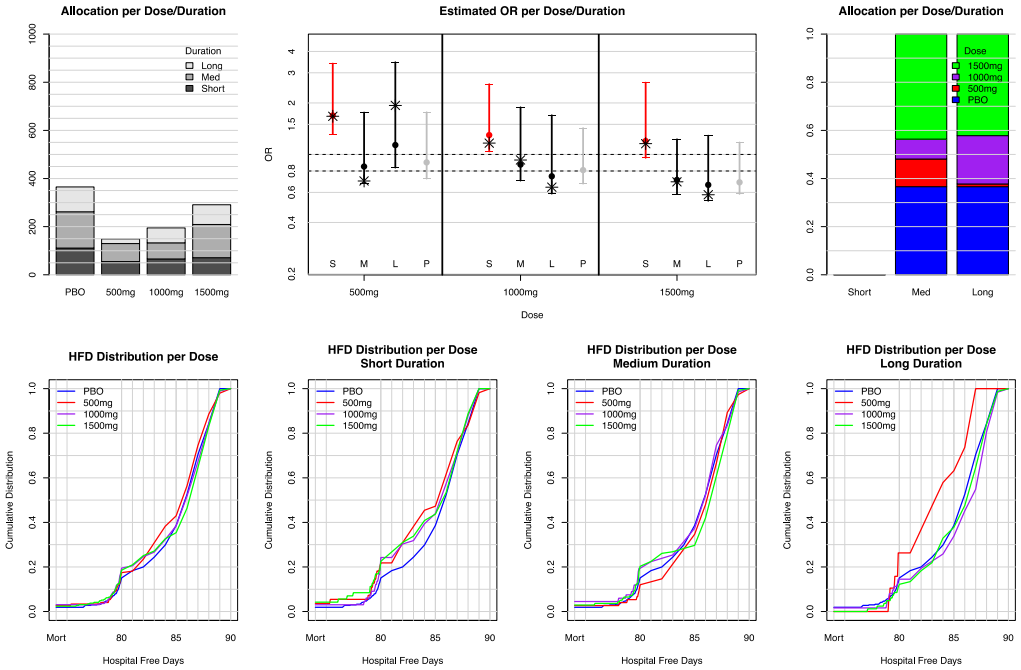

Figure 5.1.2: Example Trial 1; Interim N=1000

Figure 5.1.3 shows results from the third interim analysis when 1500 patients have 90-day data. Approximately 550 patients have been allocated to placebo, 200 to 500mg, 250 to 1000mg and 500 to 1500mg. No new patients have been enrolled in the short duration. The 500mg dose is stopped in both the intermediate and long durations. Within the intermediate and long durations, the 1500 and 1000mg doses have an OR estimated around .80 and have approximately equal allocations within each duration.

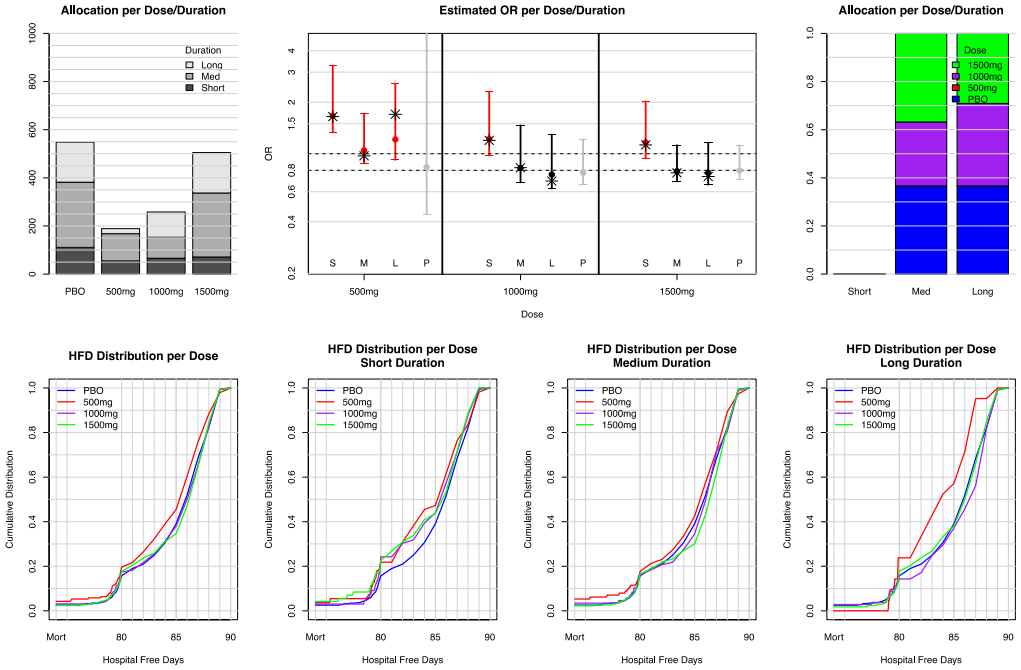

Figure 5.1.3: Example Trial 1; Interim N=1500

Figure 5.1.4 shows results from the fourth interim analysis when 2000 patients have 90-day data. Approximately 725 patients have been allocated to placebo, 200 to 500mg (no new patients), 425 to 1000mg and 675 to 1500mg. No new patients have been enrolled in the short duration. The pooled estimate across all actively enrolling durations (intermediate and long) for the 1000mg dose is approximately .75 and the upper limit of the CI has dropped below 1. Therefore, the posterior probability that the OR<1 for the 1000mg dose is greater than the interim-specific threshold (.995) and the study is stopped for success.

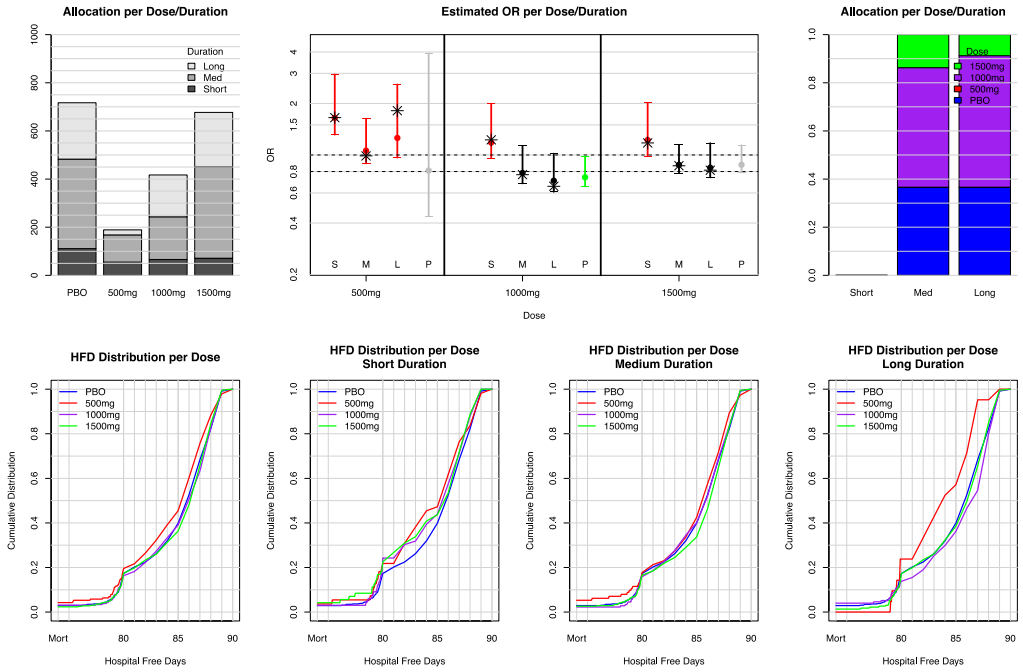

Figure 5.1.4: Example Trial 1; Interim N=2000

5.2 Example Trial 2

Figure 5.2.1 shows results from the first interim analysis when 500 patients have 90-day data. Approximately 200 patients have been allocated to placebo and 100 to 500, 1000 and 1500mg each. Within the cumulative distribution plots, the curves for each dose of metformin within each duration are mostly to the left and above the curve for placebo, indicating less HFD for each dose in each duration relative to placebo. For all doses within all durations the OR is estimated to be greater than 1.3 and the posterior probability that the OR < .8 is less than 15%. Thus, the trial stops for futility at the first interim analysis.

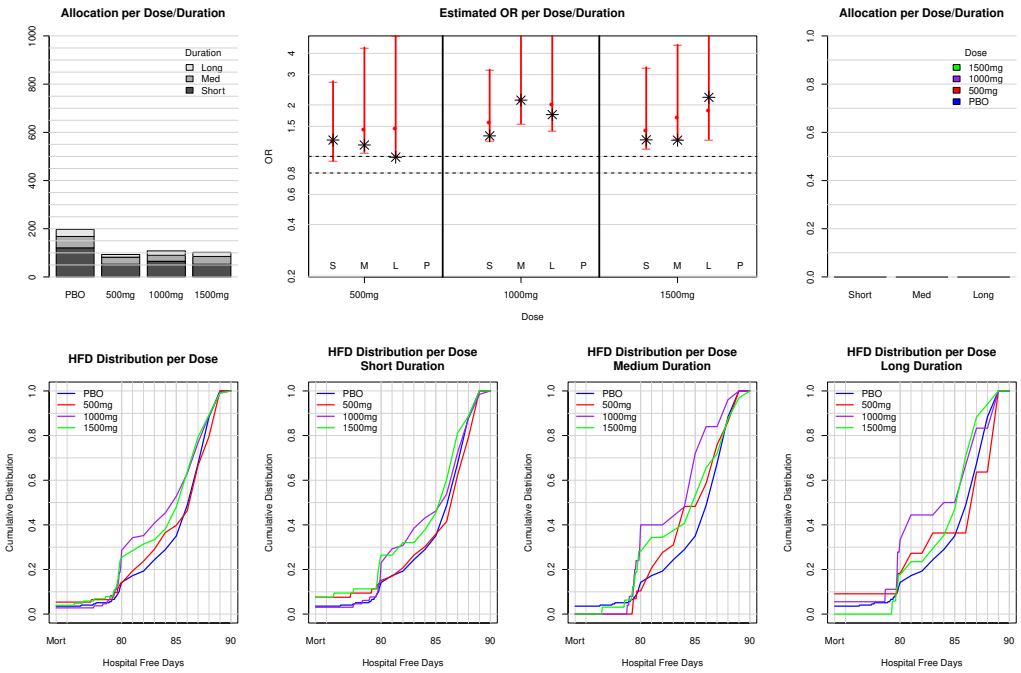

Figure 5.2.1: Example Trial 2; Interim N=500

Reitz KM, *et al. BMJ Open* 2020; 10:e037690. doi: 10.1136/bmjopen-2020-037690

6.0 Operating Characteristics

We simulate clinical trials under 7 possible treatment effect scenarios. Under the null scenario we assume that there is a 0% reduction in HD across all doses and all durations of metformin. Under all other scenarios we assume that the max effect is a 15% reduction in HD. The effect for each dose and duration is specified based on the dose-response and duration-response assumptions. We simulate under 3 different dose-response profiles. Under the "plateau" dose-response profile we assume a 7.5% reduction of the 500mg dose and a 15% reduction for the 1000 and 1500 mg doses. Under the "one good" profile we assume that there is a 0% reduction in HD for the 500 and 1000mg doses and a 15% reduction for the 1500mg dose. Under the "linear" profile, we assume a 3.75% reduction in HD for the 500mg dose, a 7.5% reduction for the 1000mg dose and a 15% reduction for the 1500mg dose. We also simulate under 2 different duration-response profiles, one where all durations work equally well and one where the intermediate and long durations work equally well but the short duration does not work for all doses. For each simulation we assume that 40% of the patients will have a short duration, 35% an intermediate duration and 25% a long duration.

Under each treatment effect scenario, we simulate 1000 clinical trials and report the following operating characteristics in Table 6.1:

- Probability of early success and total success
- Mean number of subjects enrolled in the trial
- Probability of stopping the short duration, intermediate duration or all of the durations
- Probability each dose is selected as best
- Probability increase sample size to 2500

The overall Type I error of the trial is 2.4% with 1% of the null trials stopping early for success and 91% of the null trials stopping early for futility or not increasing to the maximums sample size of 2500. The mean number of patients enrolled under the null scenario is 676. The probability the sample size is increased to 2500 under the null is 8%.

The power of the trial under the alternative scenarios ranges from 77-92% with the mean number of patients enrolled ranging from 1725 to 1822. When the short duration does not work, the probability of stopping the short duration is 80-84%. Across all alternative scenarios, we are choosing the right dose (a dose that has the maximum 15% reduction in HD effect) 79-96% of the time. Finally, the maximum sample size is increased from 2000-2500 21-31% of the time.

Table 6.1: Operating Characteristics

Dose Response	Duration Response	Prob. Success		Mean N	Prob. Stop Futility			Prob. Selected Best			Prob. Enroll 2500
		Early	Total		Short	Int.	All	500	1000	1500	
Null	**-**	0.010	0.024	676	0.95	0.92	0.91	0.35	0.30	0.35	0.08
Plateau	**All Work**	0.75	0.92	1767	0.12	0.06	0.04	0.04	0.50	0.46	0.21
	Not Short	0.66	0.87	1822	0.86	0.15	0.09	0.06	0.44	0.49	0.26
One Good	**All Work**	0.67	0.86	1729	0.28	0.15	0.11	0.03	0.03	0.95	0.23
	Not Short	0.56	0.78	1725	0.83	0.28	0.19	0.06	0.04	0.90	0.26
Linear	**All Work**	0.64	0.84	1776	0.20	0.12	0.10	0.04	0.10	0.86	0.26
	Not Short	0.51	0.77	1782	0.82	0.24	0.17	0.07	0.14	0.79	0.31