

# Tiptop statistical simulation report for the cord clamping research question: sample size calculation and operating characteristics of an adaptive design.

<i>Lead author and simulation programming</i>	<i>Munya Dimairo (<a href="mailto:m.dimairo@sheffield.ac.uk">m.dimairo@sheffield.ac.uk</a>) Clinical Trials Research Unit Sheffield Centre for Health and Related Research (SCHARR) Division of Population Health School of Medicine and Population Health</i>
<i>Contributors</i>	<i>Christopher Partlett Nottingham Clinical Trials Unit School of Medicine University of Nottingham</i>

## Table of Contents

1	Brief background and objectives .....	8
2	Trial design, trial adaptations and decision-making criteria .....	8
2.1	Trial design .....	8
2.2	The rationale for design parameters .....	8
2.3	The minimum clinically important difference.....	9
2.4	Rationale for the trial population and implication on the design.....	9
2.5	Trial adaptations and motivations .....	10
2.6	Decision-making criteria.....	10
3	Simulation methods.....	11
3.1	Choice of simulation parameters and scenarios.....	11
3.2	Approach to statistical simulations .....	13
3.3	Metrics for assessing statistical performance.....	13
4	Sample size for a fixed design.....	16
5	Simulation results for a design with one interim analysis.....	16
5.1	Impact on sample size aspects.....	16
5.1.1	Maximum sample size .....	16
5.1.2	Expected sample size. ....	17
5.1.3	The ratio of the expected sample size to the maximum and fixed design.....	18
5.2	Impact on making correct and incorrect superiority decisions. ....	20
5.3	Impact on chances of futility early stopping.....	21
5.4	Summary of simulation results for one interim analysis. ....	23
5.5	Sample sizes and operating characteristics of potential design option .....	23
6	Simulation results for a design with two interim analyses.....	26
6.1	Impact on the maximum sample size. ....	26
6.2	Impact on the expected sample size.....	27
6.3	Impact on making correct and incorrect decisions.....	28
6.4	Impact on the chances of futility early stopping. ....	29
6.4.1	The value of the first interim analysis.....	29
6.4.2	The value of the second interim analysis. ....	30
6.4.3	The value of both interim analyses.....	31
6.5	Impact on potential sample size savings.....	32
6.6	Summary of simulation results for two interim analyses. ....	34

6.7	Sample size and operating characteristics of design options .....	35
6.8	Updated sample size and operating characteristics of design options .....	41
7	Conclusions .....	46
8	References .....	46

## List of abbreviations

ARD	Absolute risk difference
DCC	Delayed cord clamping
ECC	Early cord clamping
ICC	Intracluster correlation coefficient
IDMC	Independent data monitoring committee
MCID	Minimum clinically important difference (aka targeted treatment effect)
NNRD	National Neonatal Research Database
RR	Relative risk/risk ratio
ONS	Office of National Statistics
OR	Odds Ratio

## List of Tables

Table 1. Primary outcome event rates by gestational age window. ....	9
Table 2. Targeted subgroup effects based on a consistent treatment effect of a relative reduction in death or brain injury of 35%. ....	9
Table 3. Simulation parameters and scenarios.....	12
Table 4. Metrics to assess simulations and contextual meaning. ....	14
Table 5. Sample size and operating characteristics based on 1,000,000 simulations.....	25
Table 6. Operating characteristics for interim analyses at (45%, 65%) and futility thresholds (0, 0.5) based on 500, 000 simulations. ....	37
Table 7. Operating characteristics for interim analyses at (45%, 70%) and futility thresholds (0, 0.5) based on 500, 000 simulations. ....	38
Table 8. Operating characteristics for interim analyses at (50%, 70%) and futility thresholds (0, 0.5) based on 500, 000 simulations. ....	39
Table 9. Operating characteristics for interim analyses at (50%, 75%) and futility thresholds (0, 0.5) based on 500, 000 simulations. ....	40
Table 10. Operating characteristics for interim analyses at (45%, 65%) and futility threshold (0, 0) based on 500, 000 simulations. ....	42
Table 11. Operating characteristics for interim analyses at (45%, 70%) and futility threshold (0, 0) based on 500, 000 simulations. ....	43
Table 12. Operating characteristics for interim analyses at (50%, 70%) and futility threshold (0, 0) based on 500, 000 simulations. ....	44
Table 13. Operating characteristics for interim analyses at (50%, 75%) and futility thresholds (0, 0) based on 500, 000 simulations. ....	45

## List of Figures

Figure 1. Impact on the maximum sample size. ....	17
Figure 2. Impact on expected sample size. ....	18
Figure 3. Ratio of the expected sample size to the fixed design sample size. ....	19
Figure 4. Ratio of the expected sample size to the maximum sample size. ....	20
Figure 5. Impact on making correct and incorrect superiority decisions. ....	21
Figure 6. Impact on the probability of stopping early for futility. ....	22
Figure 7. Impact on the maximum sample size. ....	27
Figure 8. Impact on the expected sample size. ....	28
Figure 9. Impact on superiority decision-making (two interim analyses). ....	29
Figure 10. The value of the first interim analysis. ....	30
Figure 11. Additional benefits of the second interim analysis. ....	31
Figure 12. The value of two interim analyses. ....	32
Figure 13. Sample size saving relative to the maximum sample size. ....	33
Figure 14. Saving in sample size relative to the fixed design. ....	34

## Executive summary

This report presents simulation results to help clinical researchers and methodologists assess the value of adaptive designs with non-binding futility early stopping options as well as to choose a feasible, efficient, and robust design to address the cord clamping research question.

Simulation results under several scenarios of the delayed cord clamping (DCC) treatment effect demonstrated that an adaptive design with one interim analysis performed with 40% to 55% of the information fraction (accrued primary outcome data) yields robust interim futility and final efficacy decisions. If one interim analysis is performed, any futility threshold critical value of 0 to 0.5 (~25.5% of the minimum clinically important difference, MCID) could be used without compromising efficacy decisions. As an exemplar, sample sizes and operating characteristics of a design with one interim analysis at 50% information fraction are presented assuming a conservative futility threshold of 0 critical value (one-sided p-value of 0.5), which gives around 51.9% probability of futility early stopping if the effect of DCC is the same as early cord clamping (ECC).

Further simulations demonstrated that conducting the second interim analysis is valuable for an adaptive design with two interim analyses allowing for non-binding futility early stopping. All combinations of information fractions where the first is performed at 40% to 50% and the second after a 20% to 25% increase in information fraction (spacing of interim analysis) are statistically useful. If the futility threshold for the first interim analysis is kept at 0 critical value, the futility threshold for the second interim analysis can be increased to around 0.6 critical value (~30.6% of the MCID) without compromising both the interim and final decisions while maximising the chances of stopping early for futility if the DCC is worse or the same as ECC. The maximum sample size increases with the increasing futility threshold of the second interim analysis and with the decreasing information fraction of the first interim analysis. As such, the chosen design should be feasible to recruit the maximum sample size.

Although an adaptive design with three interim analyses has not been assessed, it could be worthwhile to explore the value of a third interim analysis if the selected design includes a second interim analysis performed at 65% or 70% and when the futility threshold for the 2<sup>nd</sup> interim analysis is larger than at the 1<sup>st</sup> interim analysis. In addition, a very conservative futility threshold for the first interim analysis of 0 has been considered, but in theory, this could be increased depending on clinical advice and what the clinical team wants to achieve.

Finally, following discussions with the research team, the clinical team preferred an adaptive design with futility thresholds (1<sup>st</sup>, 2<sup>nd</sup>) of (0, 0) to minimise the probability of stopping early when there is a small to moderate positive benefit of DCC. As such, four updated competing design options with their sample sizes and operating characteristics are presented for the research team to choose from. The final design should be selected based on several factors such as the feasibility of recruiting the maximum sample size adjusted for dropout rate, chances of futility early stopping correctly, control of error rates, sufficient spacing between interims, potential savings in resources, and the value of conducting additional interim analyses.

## 1 Brief background and objectives

This report summarises the statistical simulation methods and results that informed the design for the cord clamping research question. It aims to guide appropriate decisions about the adaptive trial design. Specific objectives are to:

- a) explore the timing and decision rules on the statistical performance of the design under several relevant scenarios,
- b) inform the appropriate timing and decision rules for trial adaptations that result in the desired statistical properties of the adaptive design,
- c) describe the statistical performance of the chosen adaptive design accounting for trial adaptations considered under several relevant scenarios.

In principle, an efficient adaptive design is desirable to facilitate correct decision-making about the benefits of study treatment. Efficiency is contextual and may relate to savings in research resources and the ability to address research questions robustly.

## 2 Trial design, trial adaptations and decision-making criteria

To set the scene, this section briefly covers the design, primary outcome, rationale for parameter estimates that informed the design, trial population and rationale, trial adaptations and motivations behind them, and decision-making criteria.

### 2.1 Trial design

This is an open-label, two-arm, pragmatic, multi-centre, superiority, parallel-group, and group sequential (adaptive) randomised controlled trial. Eligible participants (women at 22 to 32 gestational age) will be individually randomised (1:1) to either early cord clamping (ECC) or delayed cord clamping (DCC) interventions, detailed elsewhere. The primary outcome is survival without brain injury on day 7 following delivery. Key long-term outcomes are neurodevelopment impairment at two years of age corrected for prematurity. The unit of randomisation is the mother, but the infant is the unit of analysis. Multiple births (e.g., twins) will be randomised to the same intervention allocated to the mother as informed by previous research involving engagement with patients and the public. The prevalence of multiple births from the same mother that will require resuscitation is expected to be negligible so adjusting for clustering around the mother is unnecessary. Also, existing data<sup>1</sup> suggest that the intracluster correlation coefficient (ICC) for short-term mortality and brain injury is small so likely to have a negligible impact on sample size. As such, the sample size will not be adjusted for the design effect.

### 2.2 The rationale for design parameters

Based on data from the National Neonatal Research Database (NNRD) for all admitted infants between 2016 and 2022, the background event rate of the primary outcome depends on the gestational age window and based on clinical advice (biological plausibility), the effect of DCC is likely to diminish with increasing gestational age (Table 1). That is, the treatment effect measure is a function of the background event rate and gestational age window. For example, a 4.5% absolute increase in the primary outcome event rate attributed to DCC is only plausible



in the 22<sup>+0</sup> to 27<sup>+6</sup> weeks gestational age and not in the 28<sup>+0</sup> to 31<sup>+6</sup> weeks gestational age (Table 1). As a result, the relative increase in survival without brain injury (inferred from the risk ratio/relative risk [RR]) will be the primary measure of the treatment effect of interest although the absolute risk difference (ARD) and odds ratio (OR) can be presented alongside to aid interpretation.

Table 1. Primary outcome event rates by gestational age window.

Gestational age window	Admissions	Needed stabilisation	Death or brain injury on day 7	
			n	%
22 <sup>+0</sup> to 31 <sup>+6</sup> weeks (C)	51329	32592	3438	10.5
22 <sup>+0</sup> to 27 <sup>+6</sup> weeks (A)	16492	14340	2660	18.5
28 <sup>+0</sup> to 31 <sup>+6</sup> weeks (B)	34837	18252	778	4.3

Using the Office of National Statistics (ONS) data to identify delivery room deaths suggests a further decrease in survival of around 2.3% for the 22<sup>+0</sup> to 31<sup>+6</sup> weeks resulting in the primary outcome event rate of around 87.2%.

## 2.3 The minimum clinically important difference

A 5.2% relative increase (RR of 1.052) in survival without brain injury on day 7 of delivery is viewed as clinically important to change practice. This is equivalent to a 4.5% overall absolute increase assuming a background event rate of 87.2%. This is also equivalent to a 35% relative reduction in death or brain injury. Table 2 shows the targeted subgroup effects on the ARD scale which corresponds to a 35% relative reduction in death or brain injury. For example, an 8.1% absolute increase in survival without brain injury (from 76.9%) in the 22<sup>+0</sup> to 27<sup>+6</sup> weeks gestational age is the targeted treatment effect equivalent to a 35% relative reduction in death or brain injury.

Table 2. Targeted subgroup effects based on a consistent treatment effect of a relative reduction in death or brain injury of 35%.

Gestation age window	ECC (control) event rate	DCC event rate	Relative reduction in death or brain injury	Absolute increase in survival without brain injury
22 <sup>+0</sup> to 31 <sup>+6</sup> weeks	87.2%	91.7%	35.0%	4.5%
22 <sup>+0</sup> to 27 <sup>+6</sup> weeks	76.9%	85.0%	35.0%	8.1%
28 <sup>+0</sup> to 31 <sup>+6</sup> weeks	95.7%	97.2%	35.0%	1.5%

## 2.4 Rationale for the trial population and implication on the design

The trial will enrol women at 22<sup>+0</sup> to 31<sup>+6</sup> weeks gestational age (overall population C, Table 1). Although the 22<sup>+0</sup> to 27<sup>+6</sup> weeks gestational age (subpopulation A) is expected to benefit more from DCC intervention compared to the 28<sup>+0</sup> to 31<sup>+6</sup> weeks gestational age (subpopulation B) concerning the primary outcome, the latter could also benefit more concerning long-term neurological development outcomes. This underscores the need to

enrol both subpopulations. In addition, trial recruitment is feasible when both subpopulations (A and B).

Unfortunately, recruitment to a definitive trial will be infeasible if only one subpopulation is considered, which makes an adaptive enrichment design impractical to implement though in theory, it would have been well-suited given the potential differential treatment effect on ARD in the two subpopulations (A and B). For example, one could build in an option to drop subpopulation B for futility at an interim analysis and enrich subpopulation A. However, dropping subpopulation B for futility would mean the need to increase the sample size of subpopulation B hugely to maintain a minimum high statistical power (which is infeasible given its underlying prevalence). As a result, an adaptive enrichment design was considered at the design stage, but it was not pursued further for feasibility reasons.

## 2.5 Trial adaptations and motivations

A long recruitment duration is expected involving several centres. The recruitment of participants to achieve the desired sample size for a definitive trial is expected to be challenging although it is believed to be feasible. The trial, therefore, would require substantial resources. Moreover, the primary outcome (survival without brain injury) can also be viewed as a safety outcome so there is a need to incorporate safeguards into the design to protect the welfare of participants (infants and mothers). These reasons motivated the need to incorporate formal futility analysis into the design to facilitate early stopping if DCC is potentially harmful or futile (i.e., results in worse outcomes to be viewed as unsafe or unlikely to result in substantial benefits to change practice). This will save research resources and safeguard trial participants.

From a clinical perspective, there are very small chances of early stopping for benefit because of DCC demonstrating overwhelming benefits. Early stopping for efficacy is therefore not of interest and would not be formally incorporated into the design, as doing so, would unnecessarily increase the sample size. Finally, there is little uncertainty around the sample size as prior data that informed the sample size parameters (Section 2.2) were viewed as quite robust. Thus, a formal sample size re-estimation is viewed as unnecessary although the independent data monitoring committee (IDMC) may review estimates of sample size estimates during the trial as part of their oversight responsibilities.

## 2.6 Decision-making criteria

This covers when interim analyses will be conducted and decision rules for claiming evidence at both interim and final analyses. In this context, decision rules at interim analyses relate to the level of evidence that is required to trigger early stopping for futility or harm. For example, a decision rule can be expressed as “stop for futility or harm if no relative or absolute increase in survival without brain injury (treatment effect of 0 or less) is observed at an interim analysis”. The treatment effect can be expressed in different statistical quantities that summarise the level of evidence observed (e.g., critical values, ARDs, RRs, relative increases, or p-values). One measure can be mapped one-to-one onto another measure. The futility decision rule will be non-binding in the sense that it can be overruled (when triggered at an interim analysis) for some reason without undermining or inflating the type I error rate.

When interim analyses will be conducted relates to frequency (how many times) and timing (at which points) of interim analyses. Interim analyses are resource intensive, so the number of interim analyses should be weighed against the potential gains in conducting additional interim analysis and related feasibility aspects. For this reason, in practice, most adaptive trials are rarely designed with more than 3-4 interim analyses <sup>2</sup> as the benefits diminish with more frequent interim analyses. Thus, one or two interim analyses will initially be considered in simulations. The timing is expressed as information fraction, which in this case (binary outcome), is the fraction of accrued outcome data relative to the planned sample size. The smaller the information fraction the larger the uncertainty around the treatment effect and therefore, the larger the uncertainty around trial adaptation decisions – thus, undermining trial credibility. On the other hand, whilst the longer the delay in interim analysis to increase information fraction enhances the robustness of interim decisions, it diminishes the potential benefits of trial adaptations. These trade-offs can be statistically quantified to aid decisions about appropriate decision-making criteria, which is part of this simulation work. In context, most group sequential trials are stopped early with 50% to 85% information fraction <sup>2</sup>. The median information fraction of the timing of the first interim analysis was around 40% to 65% across sectors (Qiang Zhang’s ongoing PhD research). Therefore, 40% to 55% information fraction at first interim analysis seems a reasonable timing to explore through simulations. The timing of the second interim analysis should accrue additional reasonable primary outcome data for it to be worthwhile.

### 3 Simulation methods

All statistical simulations were performed in R using the ‘rpact’ version 3.0.4 <sup>3</sup>. A large number of simulation replicates per scenario was used to achieve a very small Monte Carlo simulation error within a feasible computational time. The R simulation code is assessable via [GitHub](#). This section covers the choice of simulation parameters and scenarios, how statistical simulations were conducted, and the metrics for assessing the statistical performance of the design under specified scenarios. Based on simulation results, the sample sizes and operating characteristics of the design options are presented as examples.

#### 3.1 Choice of simulation parameters and scenarios

Table 3 summarises the simulation parameters and scenarios considered as well as the associated rationale. Of note, the information fraction here relates to the proportion of participants with accrued primary outcome data at an interim analysis relative to the required maximum sample size.

Table 3. Simulation parameters and scenarios.

Design aspect	Scenarios	Rationale
Statistical power	90% (10% type 2 error rate)	To claim benefit if the treatment works with a very high probability.
Type 1 error rate	2.5% (one-sided)	To claim benefit if the treatment does not work with a very low probability. A one-sided is considered as the direction of treatment effect is important for triggering early futility stopping.
ECC (control) event rate	87.2% survival without brain injury.	See Section 2.2.
Assumed underlying treatment effect in the DCC arm	Survival without brain injury of 85%, 86%, 87.2%, 88%, 89%, 90%, 91.7%, 92.5%, and 93.5%	To cover scenarios of the level of evidence relating to harm, no difference, small to moderate treatment effects, targeted treatment effect, and overwhelming treatment effects above the MCID
Targeted treatment effect (MCID) under $H_1$	5.2% relative increase in survival without brain injury (under $H_1$ ). A 0% relative increase (RR of 1) is assumed under $H_0$ .	See Section 2.3.
Frequency of interim analyses	One interim analysis at 40%, 45%, 50%, and 55% of the information fraction.	See Section 2.6.
	Two interim analyses (1 <sup>st</sup> , 2 <sup>nd</sup> ) at (40%, 60%), (45%, 65%), (50%, 70%), (50%, 75%), (40%, 65%), and (45%, 70%).	The spacing between interim should accrue a reasonable number of participants for additional interim analyses to be worthwhile. A 20-25% increase in information fraction is considered which is expected to generate additional data of around 300 to 500 participants.
Futility thresholds (decision rules) at interim analyses	For one interim analysis, critical values of 0, 0.1, 0.2, 0.3, 0.4, and 0.5 are considered. Note that these can be converted to other quantities such as p-value and RR.  For two interim analyses, critical value combinations (1 <sup>st</sup> , 2 <sup>nd</sup> interim analysis) of (0, 0), (0, 0.1), (0, 0.2), (0, 0.3), (0, 0.4), (0, 0.5), (0, 0.6) and (0, 0.7) are considered as justified in Section 5.5.	Indicating the low level of evidence supporting the benefit of DCC that would warrant early futility stopping. For example, if efficacy is claimed if a critical value is above 1.96, then a futility threshold of 0.5 equates to observing approx. no more than 25.5% of the targeted MCID above.
Allocation ratio	1:1	No rationale to favour one over another and this is the most optimal ratio

### 3.2 Approach to statistical simulations

The statistical simulations were conducted as follows:

- 1) Set up simulation scenarios that cover the information fraction for each interim analysis considered (e.g., at 1<sup>st</sup> interim only, at 1<sup>st</sup> and 2<sup>nd</sup> interims), futility threshold at each interim analysis, and underlying event rate in the DCC arm (Table 3);
- 2) Set the seed and the number of simulation replicates (100, 000 and 50, 000 for design with one and two interim analyses to cover 216 and 432 simulation scenarios, respectively);
- 3) Fix statistical power, type 1 error, ECC event rate, and targeted treatment effect (Table 3);
- 4) Calculate the sample size for a fixed design (no interim analysis) without continuity correction (as it is unnecessary in this case);
- 5) To start simulations:
  - set the design using “getDesignGroupSequential” for given information fractions, futility thresholds, and fixed parameters;
  - calculate the sample sizes (maximum and at each interim) to feed into the simulation using “getSampleSizeRates”;
  - simulate binary outcome data for each scenario of the underlying treatment effect using “getSimulationRates” for a set ;
  - each time record parameters and estimates of interests (e.g., futility boundaries on different scales);
  - calculate or/and record the metrics for assessing performance across all simulations.

Finally, data visualisation techniques are used to present simulation metrics such as the probability of futility early stopping and ratios of sample size (Section 3.3) across scenarios considered (futility thresholds, underlying treatment effect, and timing of interim analyses). For the final design options, 1,000,000 and 500,000 simulations were used for design with one and two interim analyses, respectively.

### 3.3 Metrics for assessing statistical performance.

Table 4 summarises the metrics that are considered in assessing the statistical behaviour of the proposed adaptive design.

Table 4. Metrics to assess simulations and contextual meaning.

Metrics	Contextual meaning	Desired properties
The maximum sample size	This is the sample size that accounts for trial adaptations and is required if the trial is not stopped early for futility (even though early futility stopping was incorporated).	The maximum sample size should be feasible and the parameters that inform it should be reliable.
The expected sample size	If we conduct a trial repeated times under the same protocol, we expect trial adaptations (e.g., futility early stopping) to be triggered in some cases and in other cases the trial would proceed without changes to reach the maximum sample size. Thus, some will stop early with a smaller sample size, and some will reach the end with a larger sample size. We get the expected sample size by averaging across all these realisations and their chances of occurring as one indicator of the long-run sample size (on average) if the trial is repeated several times.	On average, a trial that would help us reach correct conclusions with a reasonably small sample size is preferred.
The ratio of an expected sample size to that of a fixed design	Describe how the expected sample size (accounting for all possible realisations of trial adaptations over a repeated experiment) relates to the sample size of a fixed design.	In evidence regions where the treatment is viewed as futile, smaller ratios indicating savings in the sample size compared to the fixed design are preferred. On the other hand, in regions indicating benefits of treatment, this ratio should be above 1 as we expect to proceed until we reach the maximum sample size that is larger than that of the fixed design.
The ratio of the fixed sample size to the maximum sample size	Tells us how the fixed design relates to the maximum sample size of an adaptive design as expressed as a ratio. This is the penalty we pay for adapting the trial in terms of the sample size we need to commit upfront (even though we may stop early).	The inflation to the sample size of the fixed design to account for trial adaptations should be reasonable for the adaptive trial to be feasible.
The ratio of the expected sample size to the maximum sample size	Tells us how the expected sample size (accounting for all possible realisations of trial adaptations over a repeated experiment) relates to the maximum sample size.	In evidence regions where the treatment is viewed as futile, smaller ratios indicating savings in sample size are preferred. On the contrary, if the treatment is beneficial, we do not expect to stop the trial early for futility so this ratio should be close to 1.
Probability of making correct decisions	The chances of making correct decisions about the benefits of the treatment.	A good design should facilitate correct decisions most of the time. For example, if we know the treatment does not work (or is effective), then the design should help us reach that conclusion with a very high probability.
Probability of making incorrect decisions	The chances of making errors in decisions about the benefits of the treatment.	A good design should help us make small errors in our decisions. For example, if we know the treatment is

		effective (does not work), the probability of concluding that it does not work (is effective) should be very small.
Probability of stopping at interims as a result of triggered trial adaptations (futility in this case) <ul style="list-style-type: none"> <li>- at each interim</li> <li>- across all interims</li> </ul>	The chances of triggering early stopping for futility at an interim analysis and across interim analyses (where appropriate).	A good design would stop early for futility with a very high probability in evidence regions where we know the treatment does not work. Similarly, it should avoid stopping early for futility in evidence regions where we know the treatment is beneficial.

## 4 Sample size for a fixed design

Assuming an 87.2% control event rate (survival without brain injury on day 7 of delivery), a total of 1956 participants/mothers (978 per arm) will be required to preserve a 90% power and a 2.5% one-sided type I error to detect a 5.2% relative increase in this event rate (4.5% absolute difference). This assumes that no interim analyses will be performed and primary outcome data will be obtained from all randomised participants (no missing data).

## 5 Simulation results for a design with one interim analysis

This section describes the simulation results for an adaptive design with one interim analysis with a non-binding futility early stopping option. This is to assess the impact of the choice of futility threshold and timing of an interim analysis on sample size, the probability of making correct and incorrect decisions, and the probability of futility early stopping under 216 scenarios.

### 5.1 Impact on sample size aspects

#### 5.1.1 Maximum sample size

Figure 1 displays the total maximum sample size (not adjusted for dropout rate) required under different scenarios assuming the trial progresses without early futility stopping even though it was incorporated into the design (e.g., due to futility threshold not being reached or futility triggered but ignored for some reasons). The maximum sample size is calculated assuming a fixed 87.2% control event rate and an MCID of 4.5% ARD (5.2% relative increase) across all scenarios ([blue vertical line](#)). As evident, the maximum sample size increases as the futility threshold (on a critical value scale) increases from 0 to 0.5. This is a penalty paid for a potential small increase in the chances of making incorrect futility early stopping when the bar of futility evidence is lowered. On the contrary, the maximum sample size is reduced by delaying the interim analysis, which is intuitive as uncertainty in decision-making reduces with increasing interim data so the penalty paid becomes less with increasing interim information fraction. For example, assuming a futility threshold of 0% relative increase/ARD the maximum sample size is around 2008 and 1972 and when the interim analysis is performed at 40% and 50% accrual data, respectively. However, the increase in sample sizes is relatively small so can be traded off against other potential benefits such as gains in the ability to stop early for futility when DCC is not performing well (Section 5.3).

In summary, if the feasibility of recruiting the maximum sample size is a critical consideration, then large futility thresholds (e.g., critical values of above 0.3) should be avoided and delaying the interim analysis minimises the maximum sample size. However, this should be interpreted alongside other metric results described in Sections 5.2 and 5.3.



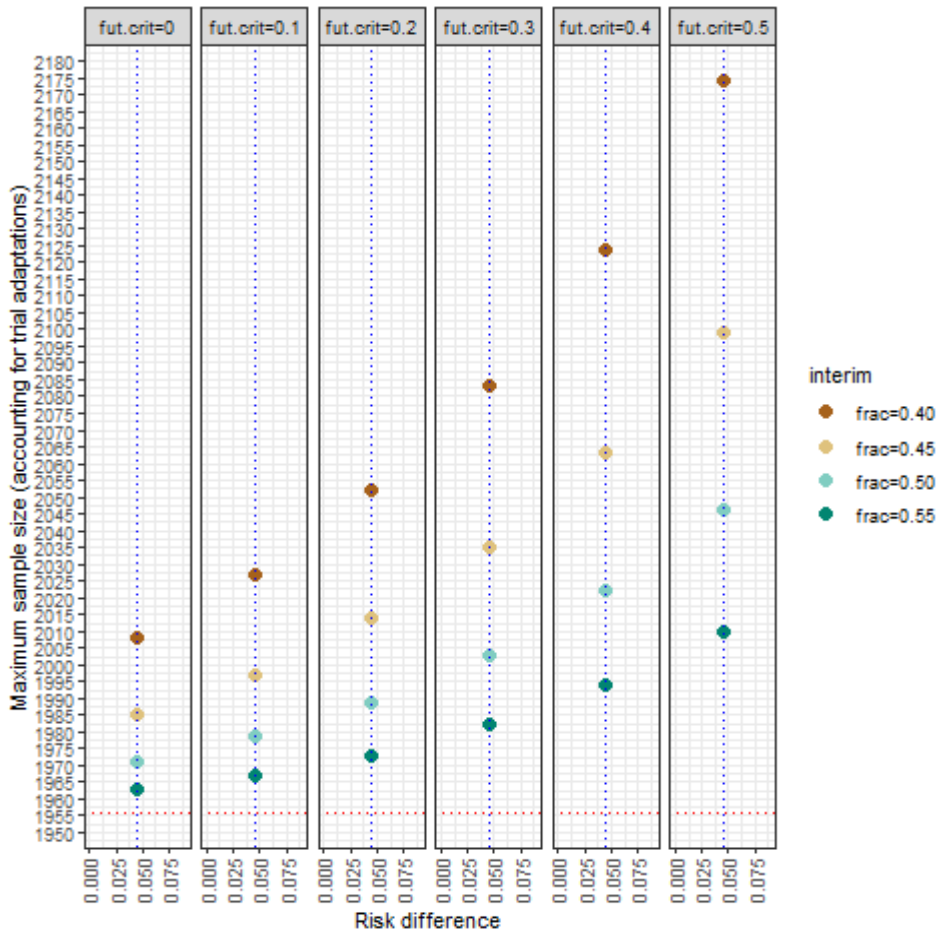


Figure 1. Impact on the maximum sample size.

### 5.1.2 Expected sample size.

Figure 2 illustrates the expected sample sizes accounting for possible realisations of trial adaptation decisions (early stopping or not) under different scenarios if the trial is conducted repeated times under a specific scenario. The timing of interim analysis has an impact on the expected sample size depending on the underlying effect of DCC. For example, if DCC is the same as or worse than ECC or the effect is very small, then conducting interim analysis earlier reduces the sample size on average. However, the opposite happens when DCC is effective. The expected sample size reduces as the futility threshold increases (i.e., being less stringent on the bar of evidence required to trigger futility early stopping). This is intuitive as the probability of stopping early increases (lowering the expected sample size) as the futility threshold increases (Section 5.3). That is, on average, one saves the sample size by increasing

the futility threshold; however, this should not be viewed in isolation from the impact on decision-making described in Sections 5.2 and 5.3.

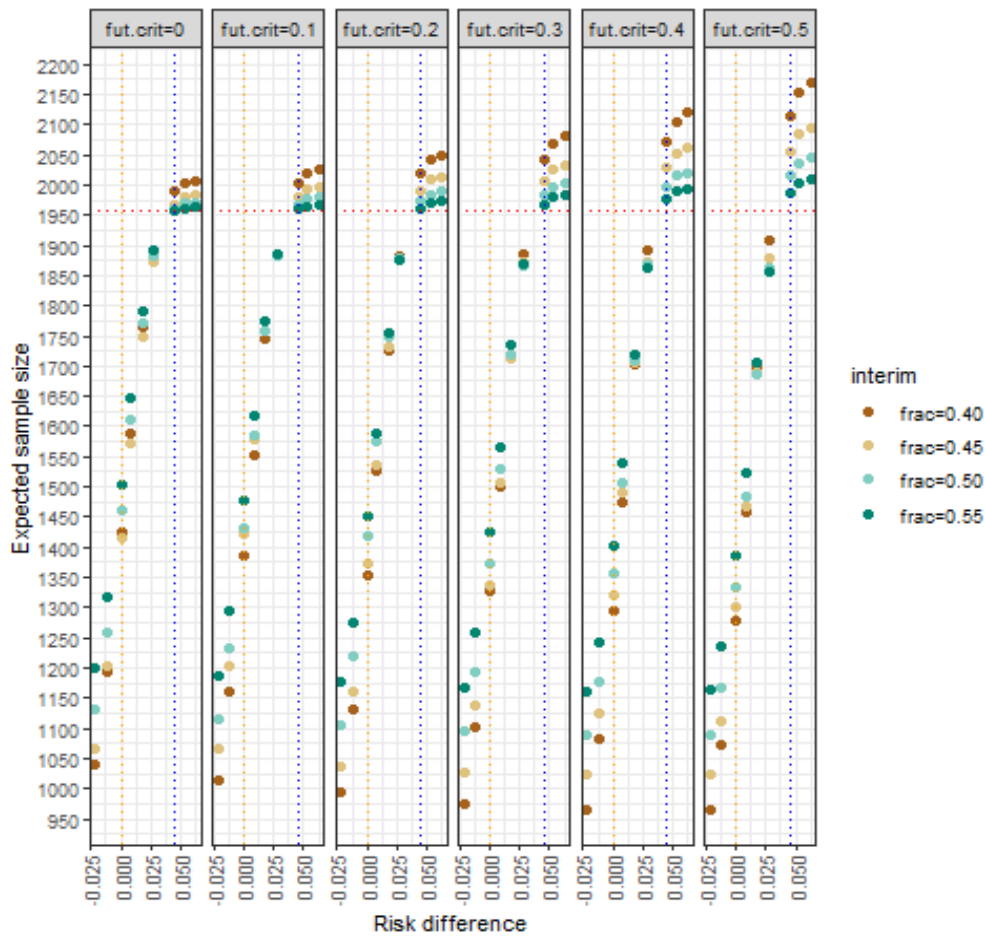


Figure 2. Impact on expected sample size.

### 5.1.3 The ratio of the expected sample size to the maximum and fixed design

The ratios of the expected sample size to the maximum sample size and the sample size of the fixed design are indicators of the average potential saving in the sample size as a result of the futility early stopping if the trial is repeated several times.

In general, on average, most sample size savings are realised when the interim analysis is performed at earlier times (i.e., 40% and 45% information fraction; Figure 3 and Figure 4). However, if DCC is effective, the trial is most likely to reach its maximum sample size (i.e., very low chances of early stopping, Figure 6) and inflation on the fixed sample size increases when the interim analysis is performed at earlier times (Figure 1). Both ratios (Figure 3 and Figure 4) are lower (if DCC is not effective) with increasing futility threshold due to increasing probability of stopping early at an interim (Section 5.3, Figure 6).

In summary, if prior signals of the efficacy of DCC are strong then one may choose to delay interim analysis as it may be unlikely to stop early and this avoids a huge penalty on the maximum sample size required if an earlier interim analysis is selected and the trial progresses

to the reach the maximum sample size. Conversely, if prior signals of the efficacy of DCC are very weak, one may choose earlier interim analysis hoping to minimise the expected sample size. Finally, the the maximum sample size is largest when the futility threshold is large – a penalty for compensating for the potential increase in type 2 error rate, however, this should be interpreted alongside other metric results (e.g., in Sections 5.2 and 5.3).

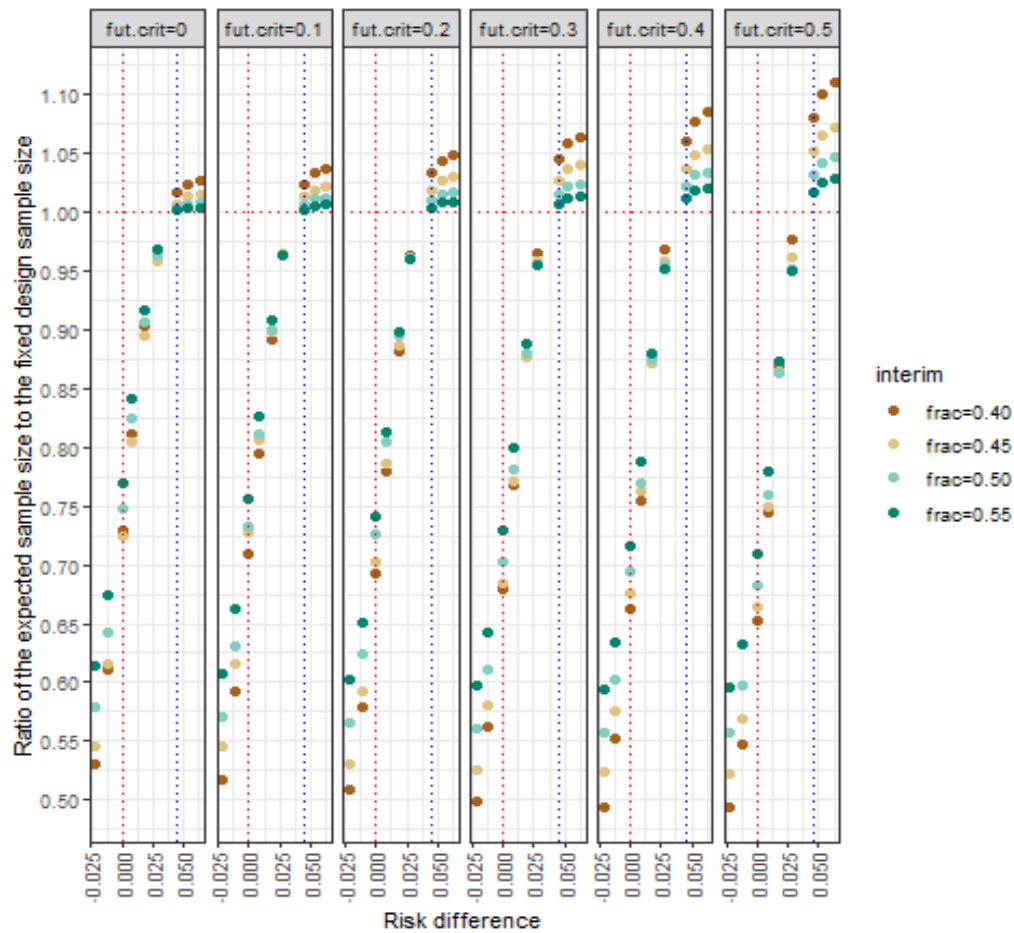


Figure 3. Ratio of the expected sample size to the fixed design sample size.

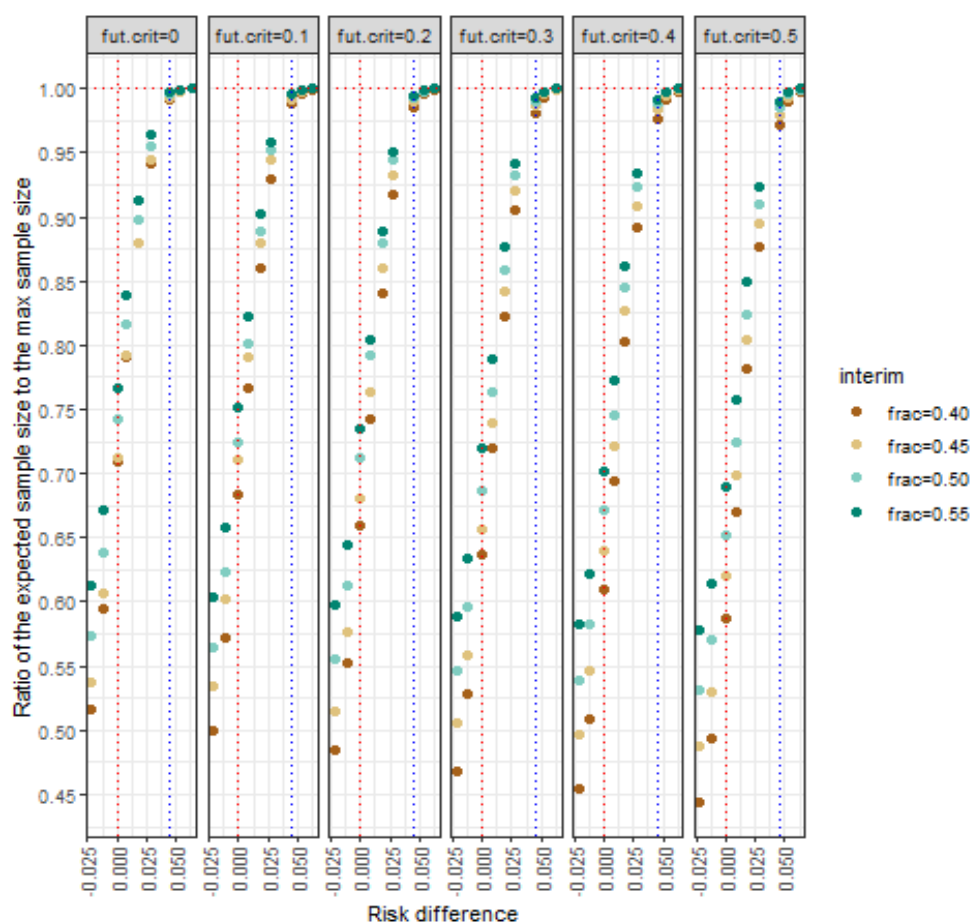


Figure 4. Ratio of the expected sample size to the maximum sample size.

## 5.2 Impact on making correct and incorrect superiority decisions.

Figure 5 shows how the probability of declaring superiority at the end of the trial changes as the treatment effect increases for different futility thresholds (on a critical value scale of 0 to 0.5) and when an interim analysis is performed as different information fractions (40% to 55%). The power corresponds to the intersection between the **green horizontal line** and the **blue vertical line**. The type 1 error rate (claiming DCC efficacy when it is not) corresponds to the intersection between the **red horizontal line** and the **orange vertical line**. As evident, the adaptive design with one futility analysis preserves the 90% power (**green horizontal line**) for a 4.5% absolute increase (5.2% relative increase, **blue vertical line**) for any of the futility thresholds considered. Similarly, the overall one-sided type 1 error rate is maintained as planned at 2.5% (**red horizontal line**) for a 0% absolute increase or relative increase (**orange vertical line**) regardless of the timing of interim analysis and futility threshold considered. This is expected as the type 1 error is expected as the sample size for each of these scenarios is calibrated to ensure that this is achieved (Section 5.1).

In summary, any combination of the timing of interim analysis and futility threshold does not compromise the final decisions about the efficacy of DCC. However, this should be interpreted alongside other metrics such as impact on maximum or expected sample sizes (Sections 5.1 and 5.3) probability of futility early stopping should be considered.

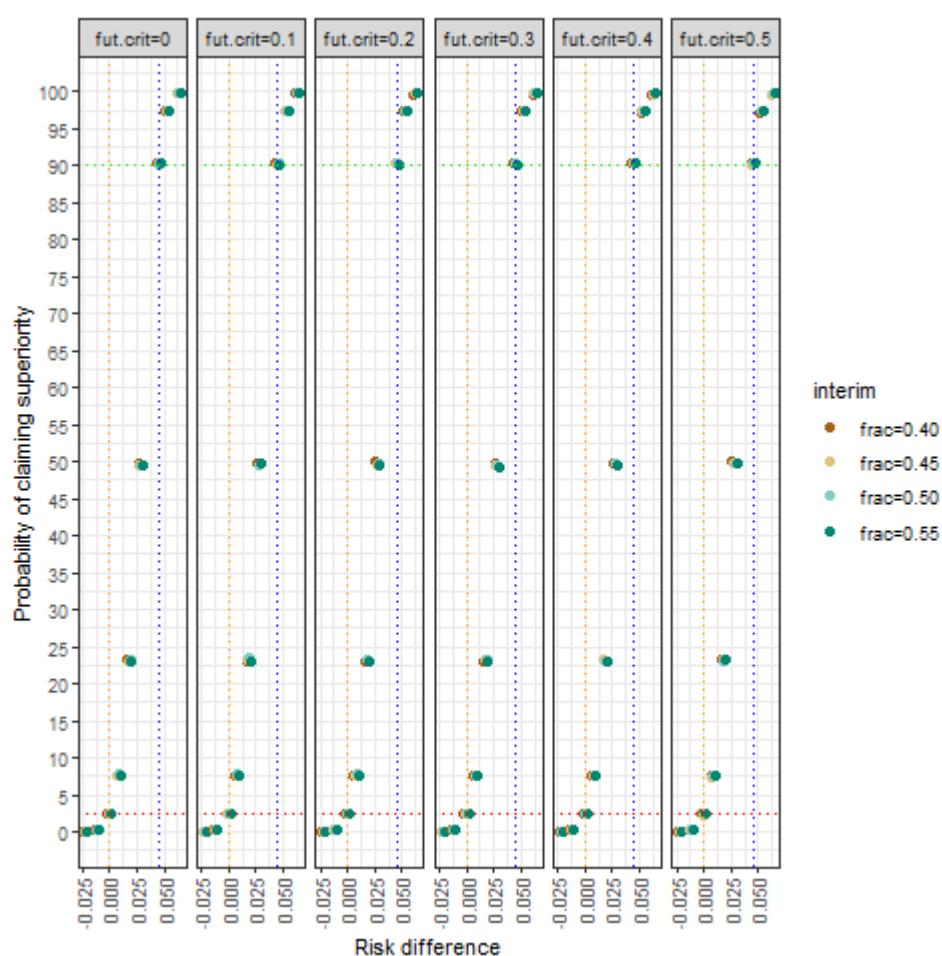


Figure 5. Impact on making correct and incorrect superiority decisions.

### 5.3 Impact on chances of futility early stopping

Figure 6 shows the probabilities of stopping for futility when a specific futility threshold (on a critical value scale of 0 to 0.5) is used at an interim analysis corresponding to an information fraction of 40% to 55% as the underlying DCC treatment effect changes (from worse to beneficial).

First, as the futility threshold increases, the probability of early stopping increases regardless of the timing of interim analysis. If the effect of DCC is the same as ECC (0% ARD, **orange vertical line**), as the futility threshold increases from 0 to 0.5 critical value, the smallest probability of futility early stopping increases across all interims from approximately 47.5% to 68%, respectively, and the probability is even much higher if DCC is worse than ECC (left region of the **orange vertical line**). Although the chances of incorrectly stopping early for futility when the targeted treatment is observed (**blue vertical line**) increase slightly, as the futility threshold increases, especially for critical values above 0.4 and when the interim analysis is performed earlier, this is not concerning as the overall type 2 error (inferred from power) is controlled (Figure 5).

Second, the impact of the timing of interim analysis is more apparent when the effect of DCC is worse than or the same as ECC, with the least probability of stopping observed when the interim analysis is performed at 40%. In general, for smaller futility thresholds, the probability of early stopping for futility is maximised when the interim analysis is delayed (i.e., at 50% or 55%). Interim analysis at an earlier time only results in the largest probability of stopping for moderate treatment effects below the MCID. Finally, larger futility thresholds are associated with large probabilities of stopping when the effect of DCC is moderate or close to the MCID.

In summary, any futility threshold critical value from 0 to 0.5 (~25.5% of the MCID) can be used; this corresponds to observing a one-sided p-value of 0.5 to 0.3085 or a relative increase of 0% to 1.13%, respectively. However, if one is very conservative and interested in minimising the chances of futility stopping early when the effect of DCC is moderate or close to the MCID, then large critical values (e.g., above 0.4) could be avoided. In addition, interim analysis at 45% to 55% seems reasonable as they give similar performance concerning the probability of futility early stopping. However, this should be interpreted alongside the feasibility of recruiting the maximum sample size (Figure 1).

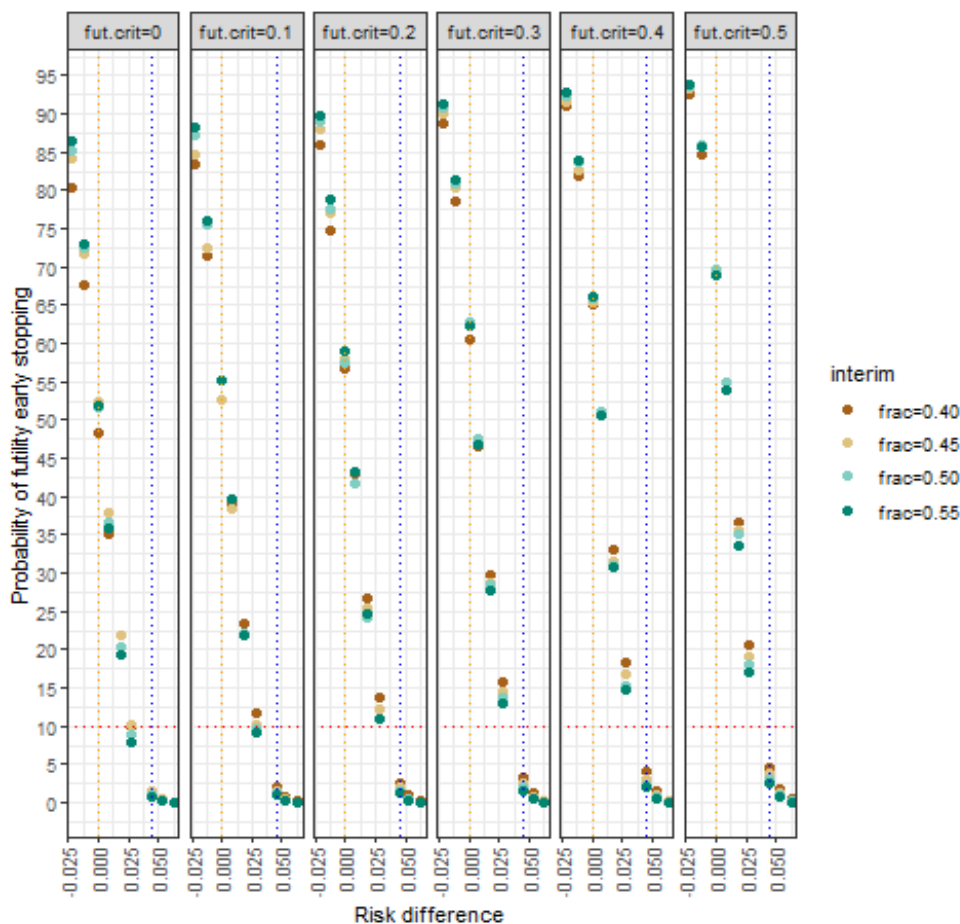


Figure 6. Impact on the probability of stopping early for futility.

## 5.4 Summary of simulation results for one interim analysis.

The simulation results demonstrated that one interim analysis for non-binding futility is valuable and it should at least be considered. Section 6 evaluates the value of the design with two interim analyses. If only one interim analysis with futility early stopping is planned, the following should be considered:

- a) all futility threshold critical values between 0 and 0.5 inclusive result in robust interim and final analyses without compromising decision errors so they can be confidently used;
- b) if one is very conservative and wants to minimise the chance of early stopping when the DCC treatment effect is moderate, then a lower futility threshold close to 0 should be used;
- c) any interim analysis with an information fraction between 40% to 55% is valuable and yields robust interim futility decisions;
- d) interim analysis conducted earlier (e.g., at 40%) requires a larger maximum sample size (Figure 1) so this should be weighed against feasibility;
- e) when the futility threshold is small (close to 0), the probability of early stopping is maximised when interim analysis is performed at later times; however, this diminishes with increasing futility threshold;
- f) a second interim analysis is required to increase the chances of futility early stopping further if DCC is futile or harmful so Section 6 explores this option.

If a second interim analysis is considered and a futility threshold of 0 or close to zero is used for the first interim analysis, then the futility threshold for the second interim analysis can be increased as more data are accrued to improve the chances of futility early stopping while not compromising efficacy decisions. This suggests that the following scenarios are worth exploring via simulations:

- Futility threshold critical values at interims (1<sup>st</sup>, 2<sup>nd</sup>): (0, 0), (0, 0.1), (0, 0.2), (0, 0.3), (0, 0.4), (0, 0.5), (0, 0.6), and (0, 0.7)
- Interim analysis information fraction (1<sup>st</sup>, 2<sup>nd</sup>): (0.40, 0.60), (0.40, 0.65), (0.45, 0.65), (0.45, 0.70), (0.50, 0.70), and (0.50, 0.75)

## 5.5 Sample sizes and operating characteristics of potential design option

Assuming an 87.2% control event rate, a maximum total of 1972 participants/mothers (986 per arm) will be required to preserve a 90% power and a 2.5% one-sided type I error to detect a 5.2% relative increase in this event rate (4.5% absolute difference). This assumes a one non-binding futility interim analysis at 50% accrual (493 per arm with primary outcome data). A trial will be stopped early for futility or harm if DCC is worse or similar to ECC; equivalent to observing a one-sided p-value of at least 0.5 or an ARD of no more than 0% ( $RR \leq 1$ ). No early stopping for efficacy is allowed. The superiority of DCC will be claimed if the critical value at the end of the trial is above 1.96 or a one-sided p-value of less than 0.025 is observed.

Table 5 details the statistical performance of the adaptive design. For example, if the two treatments are similar (0% RD, row in red), there is a 51.9% chance of stopping early for

futility. The expected sample size is 1460 accounting for the fact that the trial would proceed to the end 48.1% of the time and the ratio of this expected sample size to the maximum sample size and sample size for the fixed design is 0.741 and 0.747, respectively. On the other hand, if the targeted treatment effect is observed (row in green), there is only a negligible 1.2% chance of stopping early for futility and a 90.2% power (as expected). As the effect of DCC increases above the targeted effect (e.g., row in orange), the probability of stopping early for futility approaches 0. If DCC is harmful (e.g., worse than ECC by 1.2% ARD), the probability of stopping early increases to 72.6%.

The sample sizes presented here are for an individually randomised controlled trial without accounting for other factors such as clustering (i.e., no clustering), dropout rate (i.e., assuming 0%), and adherence (i.e., assuming 0%). If any of these factors need to be accounted for, then the sample sizes presented here (at interim and final analyses) should be inflated accordingly. For example, if clustering is an issue, then inflated sample sizes can be obtained by multiplying the interim and final sample sizes by an appropriate design effect such that interim analysis is performed when the design effect inflated interim sample size is achieved.



Table 5. Sample size and operating characteristics based on 1,000,000 simulations.

Interim Inf Frac <sup>1</sup>	DCC	ECC	RD	RR	Futility threshold			Max total SS	Total SS at interim analysis	Efficacy crit value <sup>3</sup>	Probability of futility early stopping	Statistical power	Expected sample Size	Ratio of expected SS to:	
					Critical value	RR	p value <sup>2</sup>							max SS	fixed design SS
0.5	85.0%	87.2%	-0.022	0.9748	0	1	0.500	1971	986	1.96	85.21%	0.04%	1132	0.574	0.579
0.5	86.0%	87.2%	-0.012	0.9862	0	1	0.500	1971	986	1.96	72.59%	0.30%	1257	0.637	0.643
0.5	87.2%	87.2%	0.000	1.0000	0	1	0.500	1971	986	1.96	51.91%	2.47%	1460	0.741	0.747
0.5	88.0%	87.2%	0.008	1.0092	0	1	0.500	1971	986	1.96	36.96%	7.65%	1607	0.815	0.822
0.5	89.0%	87.2%	0.018	1.0206	0	1	0.500	1971	986	1.96	20.49%	23.12%	1770	0.898	0.905
0.5	90.0%	87.2%	0.028	1.0321	0	1	0.500	1971	986	1.96	9.06%	49.46%	1882	0.955	0.963
0.5	91.7%	87.2%	0.045	1.0516	0	1	0.500	1971	986	1.96	1.19%	90.17%	1960	0.994	1.002
0.5	92.5%	87.2%	0.053	1.0608	0	1	0.500	1971	986	1.96	0.32%	97.41%	1968	0.998	1.007
0.5	93.5%	87.2%	0.063	1.0722	0	1	0.500	1971	986	1.96	0.04%	99.75%	1971	1	1.008

<sup>1</sup> information fraction at an interim analysis; <sup>2</sup> one-sided p-values; <sup>3</sup> critical value thresholds for claiming superiority at the end of the trial, DCC, delayed cord clamping (event rate); ECC, early cord clamping (event rate); max, maximum; RD, risk difference; RR, risk ratio/relative risk; SS, sample size; fixed design sample size = 1956; No efficacy early stopping is allowed. The maximum sample size should be rounded upwards to the nearest event number.

If we account for a 5% dropout rate, the maximum sample size is 2076 (1038 per arm).

## 6 Simulation results for a design with two interim analyses.

This section describes simulation results for an adaptive design with two interim analyses for a non-binding futility early stopping. This is to evaluate the value of the second interim analysis as well as the impact of the choice of futility thresholds and timing of interim analyses on sample size, the probability of making correct and incorrect decisions, and the probability of futility early stopping under 432 scenarios considered.

### 6.1 Impact on the maximum sample size.

Trial designs with the first interim analysis performed earlier require a slightly larger maximum sample size than those with a delayed first interim analysis (Figure 7). Designs with the first interim analysis conducted at the same time give similar maximum sample sizes, especially when the second futility threshold is less than 0.5 critical value. Increasing the futility threshold for the second interim analysis while keeping the futility threshold for the first interim analysis constant increases the maximum sample size. This is the penalty for a potential increase in chances of making incorrect futility decisions (type 2 error) as the futility threshold increases.

If keeping the maximum sample size smaller is critical, then one should consider the timing of interim analysis combinations with a delayed first interim analysis (e.g., excluding 40%). These results should be interpreted alongside the results of other metrics such as the impact on decision-making (Section 6.3), chances of early stopping (Section 6.4), and savings in sample size (Section 6.5).

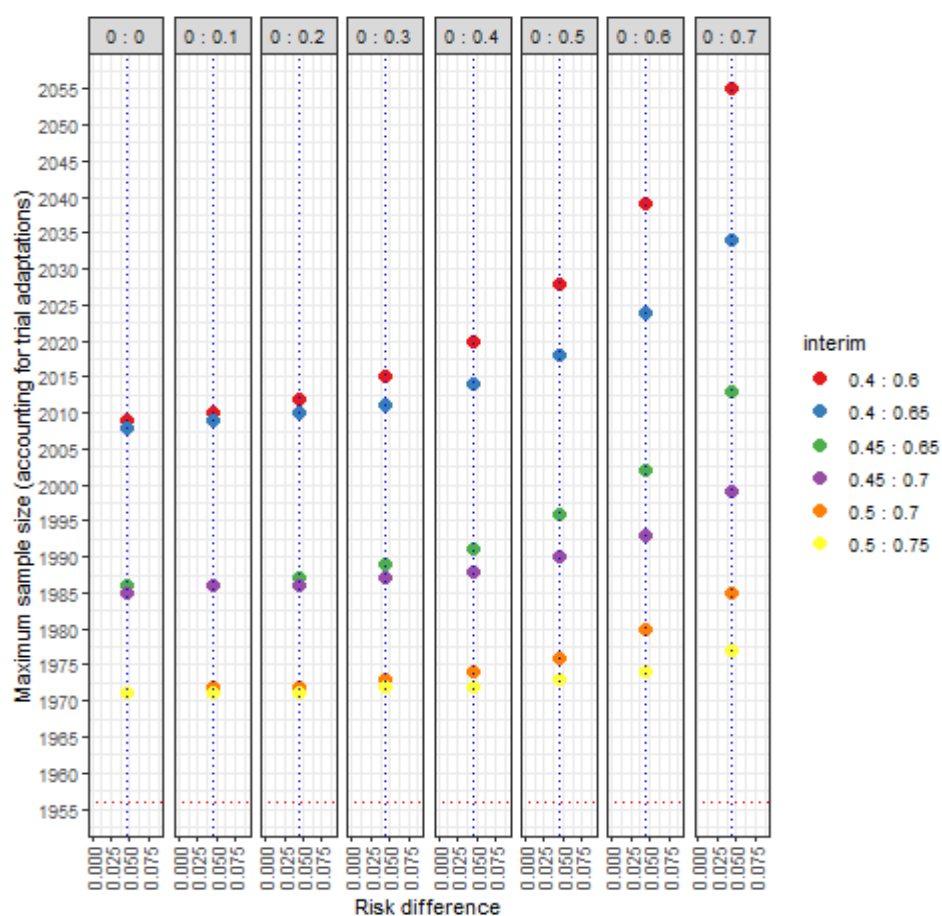


Figure 7. Impact on the maximum sample size.

## 6.2 Impact on the expected sample size.

If DCC is effective (above the [blue vertical line](#)), the expected sample size is largest when the first interim analysis is performed earlier, especially at 40%. On the other hand, if the effect of DCC is very small, the same as ECC or worse than ECC, the expected sample size is minimised by designs with the first interim analysis performed earlier, specifically at 40%. As the effect of DCC gets closer to the targeted treatment effect, all the design options yield comparable expected sample sizes, but larger when the first and second interim analyses are delayed. The choice of a futility threshold between 0 and 0.7 for the second interim analysis has the effect of reducing the expected sample size as the futility threshold increases. This is because of the increasing probability of early stopping as the futility threshold increases (Section 6.4.3 and Figure 12).

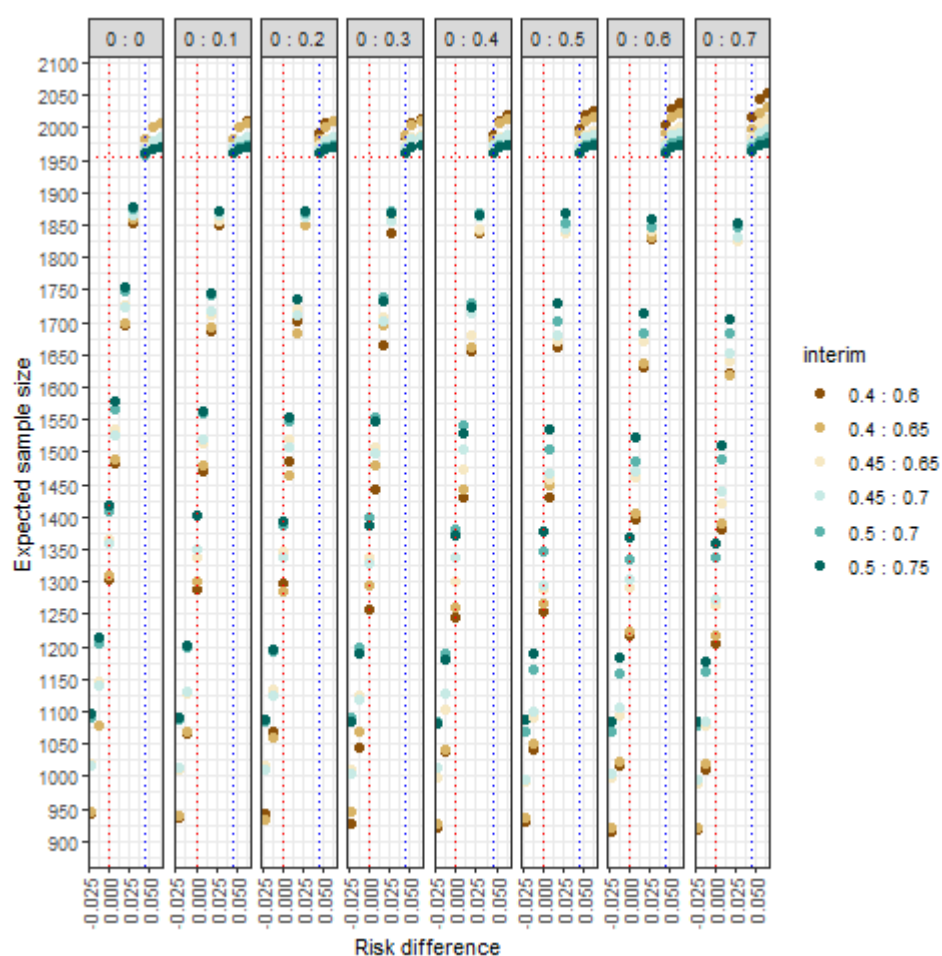


Figure 8. Impact on the expected sample size.

### 6.3 Impact on making correct and incorrect decisions.

The timing of the two interim analyses and futility threshold combinations considered do not impact the statistical power (intersection of the **blue vertical line** and **green horizontal line**) and one-sided type I error rate (intersection of the **red horizontal line** and **orange vertical line**) (Figure 9). However, there is a very negligible small loss in power when the second futility threshold for the second interim analysis is 0.7 critical value (~35.7% of the targeted effect size).

In summary, all these design options result in similar and robust efficacy decisions of DCC at the end of the trial. However, if one is very conservative for strict control of both type 1 and 2 error rates, a futility threshold region of at least 0.7 critical value should be avoided for the second interim analysis.

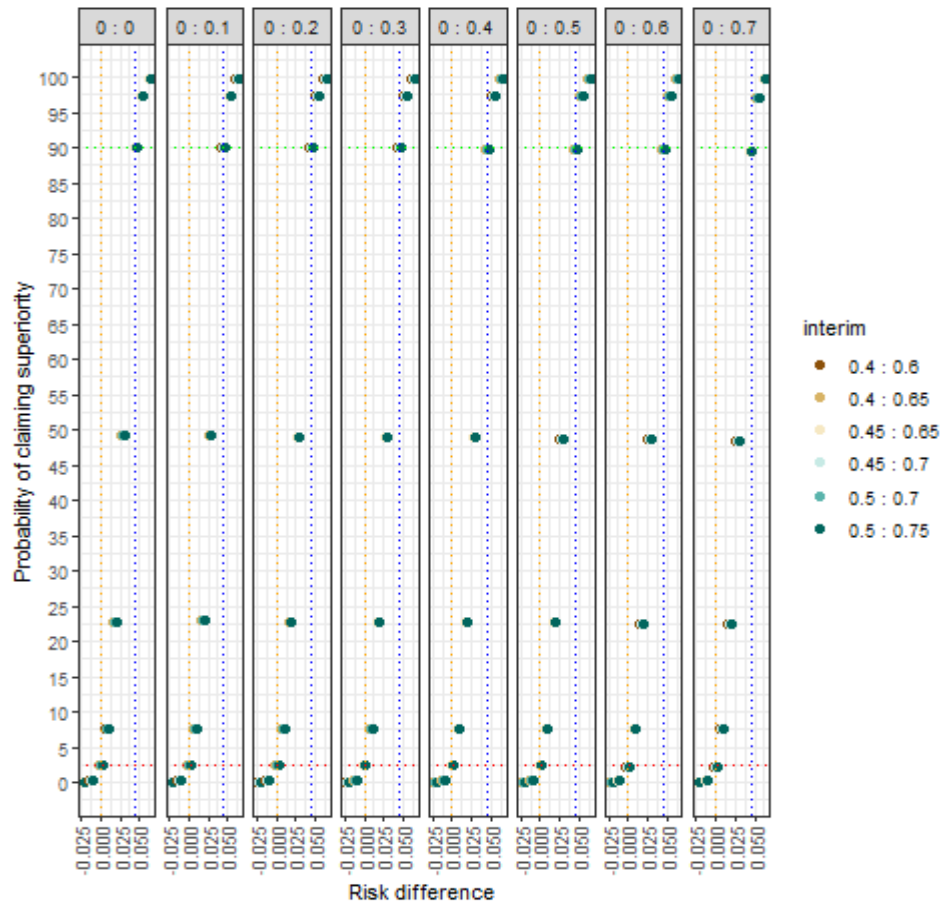


Figure 9. Impact on superiority decision-making (two interim analyses).

## 6.4 Impact on the chances of futility early stopping.

### 6.4.1 The value of the first interim analysis.

If the effect of DCC is the same as ECC (**orange vertical line**) the probability of stopping at the first interim analysis ranges from approximately 47% to 52.5% across all design options considered (Figure 10), which is essentially similar to Figure 6 (when futility threshold is zero). This probability reaches around 85% if DCC is worse than ECC by 2.2%. Results are similar for designs with a second futility threshold of 0 and 0.1; however, some differences are apparent (but within the 5% margin) as the second futility threshold increases to 0.7 critical value. In summary, if one is only interested in maximising the chance of futility stopping at the first interim analysis, these design options seem comparable with subtle differences.

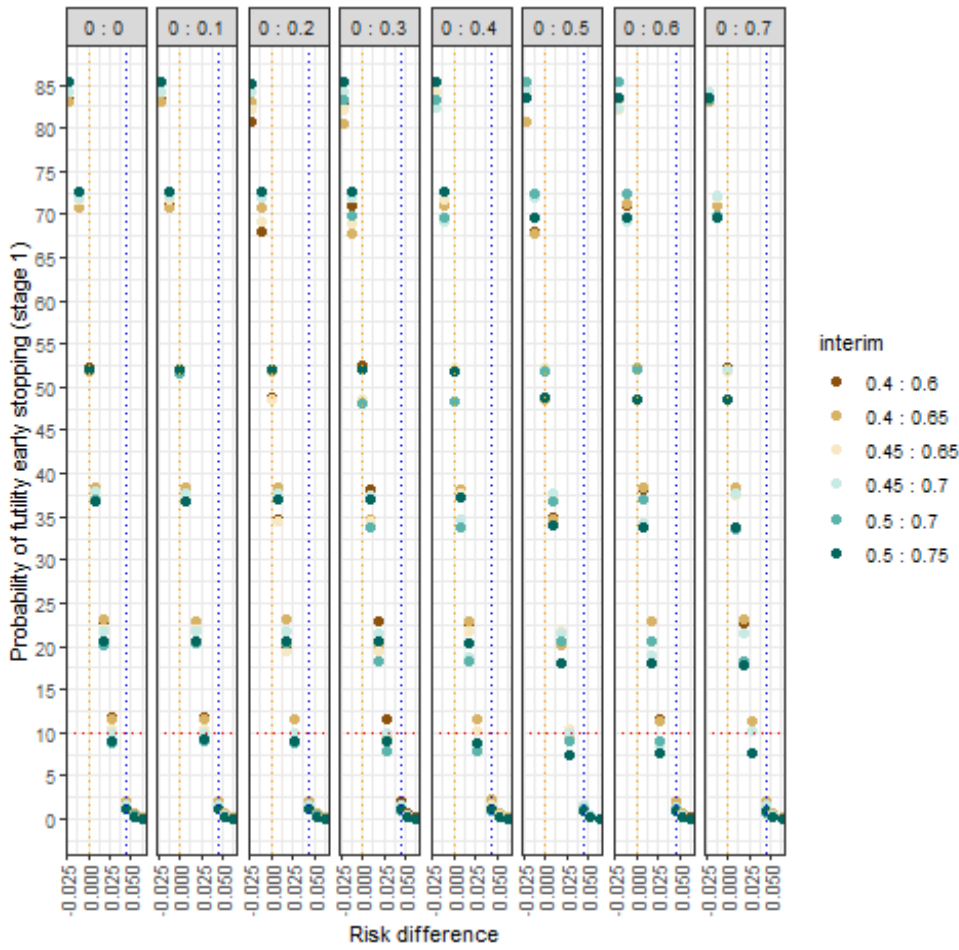


Figure 10. The value of the first interim analysis.

#### 6.4.2 The value of the second interim analysis.

Figure 11 displays the probability of stopping at the second interim analysis conditional on the trial passing the first interim analysis. It is, therefore, expected that the number of trials that pass the first interim analysis to increase as the treatment effect increases and vice versa. As such, if DCC is worse than or the same as ECC, most trials will be stopped early at the first interim analysis (Figure 10) – thus, in this region, the probability of stopping early at the second interim analysis will be small because there are fewer such trials at the second interim analysis. However, what is evident is that the probability of early stopping for futility increases drastically as the futility threshold increases. For example, if DCC is the same as ECC, this probability of futility stopping at the second interim analysis increases from approximately 11% to 28% when the futility threshold critical value is increased from 0 to 0.7 (~35.7% of the targeted treatment effect).

In summary, a second interim analysis performed at any of the information fractions considered is worthwhile and its value is maximised by increasing the futility threshold at the second interim analysis while not compromising efficacy decisions (Section 6.3).

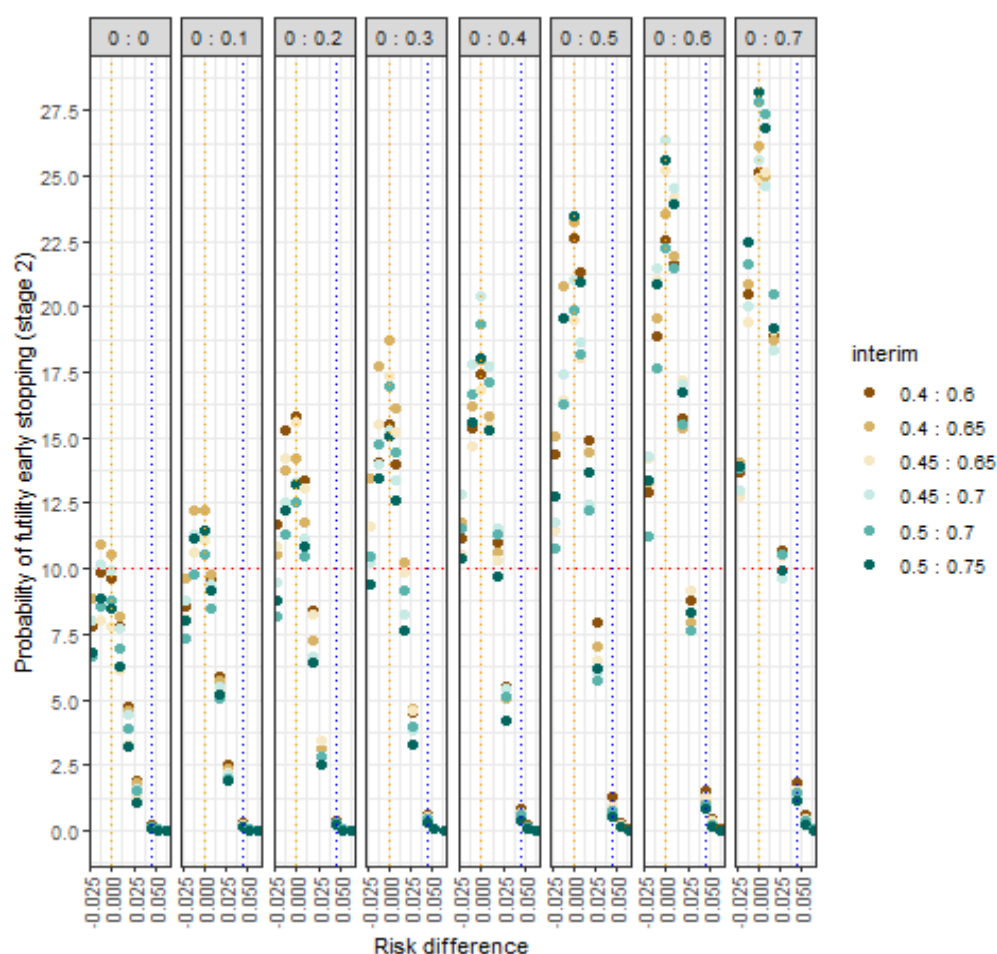


Figure 11. Additional benefits of the second interim analysis.

### 6.4.3 The value of both interim analyses.

Figure 11 displays the overall futility stopping across interim analyses. This is essentially, the sum of probabilities of stopping for futility either at the first (Section 6.4.1) or second interim analysis (Section 6.4.2). First, within the scenarios considered, the timing of the first and second interim analyses has a small impact on the overall probability of futility early stopping. Specifically, if DCC is the same as ECC, the probability of early stopping is very similar across combinations of timing of interim analyses. Small differences occur when the effect of DCC is slightly smaller than the target treatment effect and in such a region, a design that minimises early stopping may be preferable (e.g., those with delayed first interim analysis at 45% or 50%).

As the futility threshold for the second interim analysis increases from 0 to 0.7 critical value, the overall probability of futility early stopping also increases from approximately 60% to 77.5%, respectively, when the effect of DCC is the same as ECC. Of note, when the DCC treatment effect is as targeted (5.2% relative increase or 4.5% ARD) or more, there is a very small chance of incorrectly stopping early for futility across all design options. Note that this is a partial type 2 error as some trials that progress beyond the second interim analysis will be rejected at the final test. Thus, the overall type 1 error should be inferred from Figure 9.

In summary, designs with large futility thresholds for the second interim analysis are preferable. Trial designs with the combination of the timing interim analyses options considered are comparable and competing so any one of these can be chosen (based on the probability of early stopping alone). However, if one is interested in minimising the chances of early stopping when the effect of DCC is close to but less than the targeted treatment effect, designs with first interim analysis at 40% could be avoided (but these differences can be deemed very small to be of material importance).

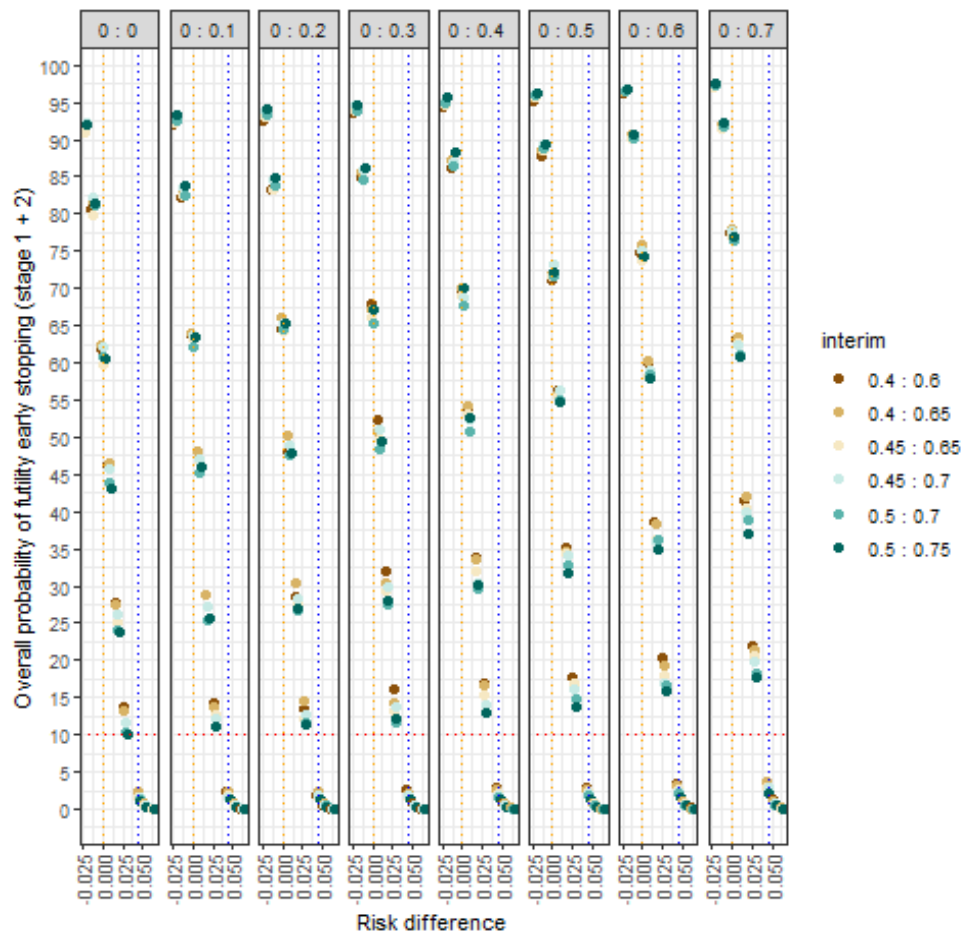


Figure 12. The value of two interim analyses.

## 6.5 Impact on potential sample size savings.

Figure 13 should be interpreted alongside results presented in Figure 7 (on maximum sample size) and Figure 12 (on the overall chances of futility early stopping). The average savings in the sample size relative to the maximum sample size increases (ratio decreases) with increasing futility threshold for the second interim analysis. Moreover, this average saving is maximised with designs with interim analysis performed earlier. However, this is expected as these designs require a large maximum sample size (Figure 7) but their probabilities of futility early stopping are similar to other design options (Figure 12). As such, a fairer comparison is shown in Figure 14, which displays sample size savings relative to the fixed design sample size (Section 4, which is constant across all scenarios).



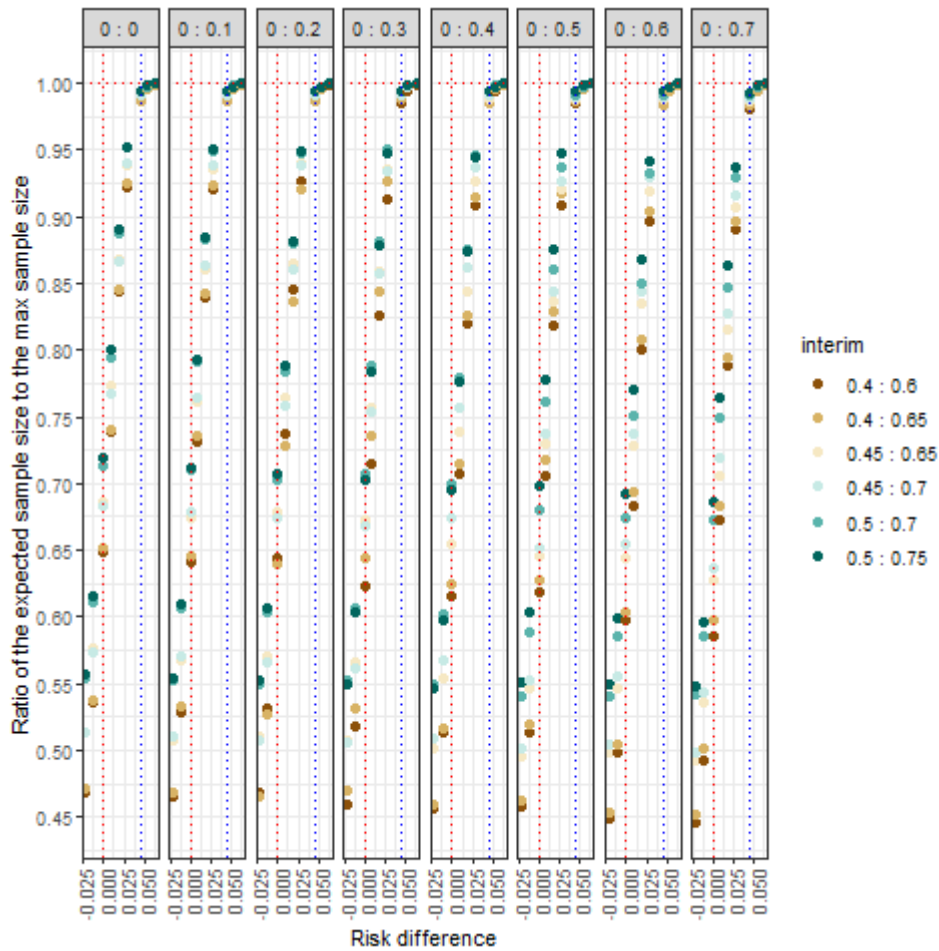


Figure 13. Sample size saving relative to the maximum sample size.

As shown in Figure 14, the average sample size savings relative to the sample size of the fixed design increase (i.e., the ratio decreases) as the futility threshold for the second interim analysis increases from 0 to 0.7 critical value. Moreover, this average sample size saving is maximised with designs that perform the first interim analysis earlier (at 40% with 45% as the middle ground). That is, across all scenarios, the average sample size saving is smaller when the first interim analysis is performed at 50%. These results should be interpreted alongside other results presented in Section 6.1 (on maximum sample size), Section 6.3 (impact on efficacy decisions), and Section 6.4.3 (probability of futility early stopping).

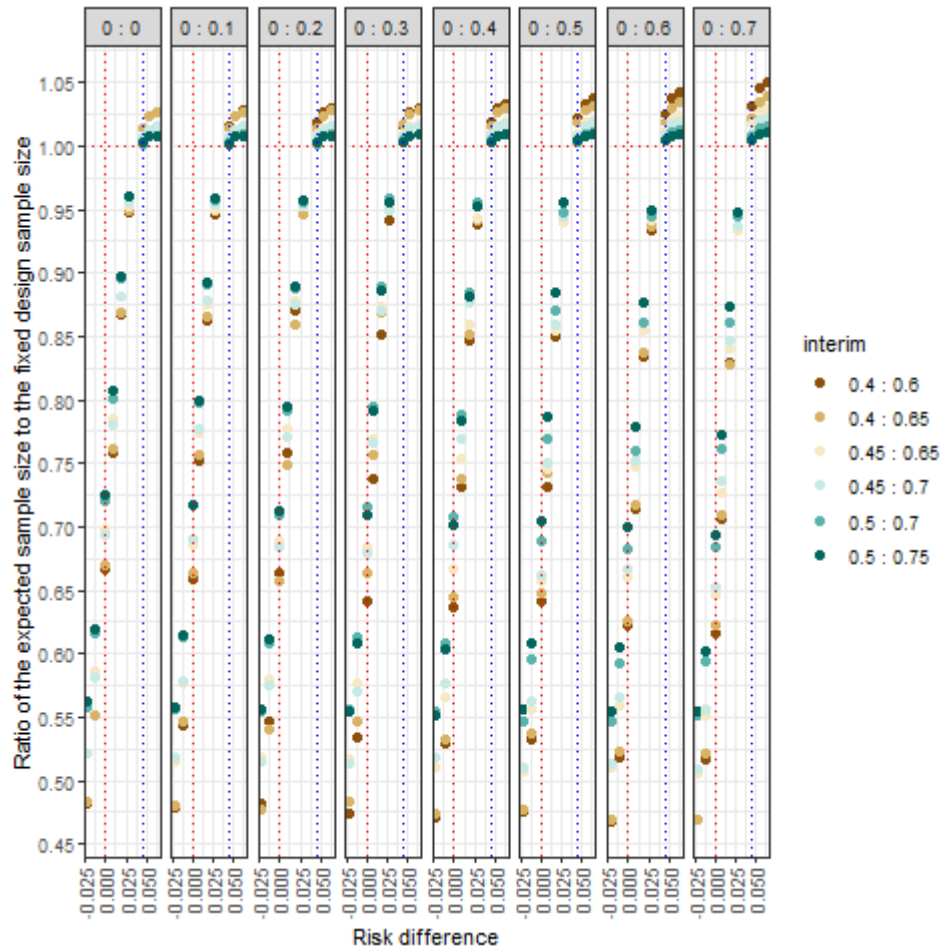


Figure 14. Saving in sample size relative to the fixed design.

## 6.6 Summary of simulation results for two interim analyses.

The choice of an appropriate design involves weighing up competing factors that include the maximum sample size required and feasibility of recruitment, impact on making robust efficacy decisions, chances of futility early stopping, and what researchers want to achieve in a particular context that requires clinical and methodological judgements. However, the following conclusions can be drawn from the simulation results:

- 1) the second interim analysis is hugely valuable as it increases the probability of early stopping by around 11% to 28% (as the futility threshold critical value is increased from 0 to 0.7) when the effect of DCC is the same as ECC;

- 2) the futility threshold for the second interim analysis should be increased and critical values as large as 0.7 can be considered. However, if one is very conservative, a critical value of at least 0.7 (~35.7% of the targeted treatment effect) could be avoided. On this basis, from a statistical perspective, a futility threshold value of below 0.7 (e.g., 0.6 or 0.65 critical value scale) can be used for the second interim analysis;
- 3) any one of the information fraction combinations for the interim analyses considered leads to robust interim and final decisions;
- 4) when the first interim analysis is performed earlier, the maximum sample size increases and this increase becomes larger as the second interim analysis futility threshold increases. However, some of these increases can be viewed as relatively small in the absolute number of participants so recruitment feasibility should be considered;
- 5) if one is interested in minimising the chances of stopping early when the effect of DCC is close to but less than the MCID, then designs that perform interim analysis earlier could be avoided (but differences can be viewed as very small), and;
- 6) the perceptions about the underlying treatment effect (e.g., based on prior signals of DCC efficacy) can help choose suitable designs as some metrics depend on assumptions about this underlying treatment effect.

The value of a third interim analysis may need to be explored, especially if the second interim analysis is performed approximately below 70% of the information fraction and the futility threshold for the second interim analysis is larger than the one used for the first interim analysis. For example, at 65% information fraction for the second interim analysis, a third interim analysis at 85% information fraction is worth considering as it could lead to a sample size saving of approximately 250 to 300 participants if the trial is stopped early for futility at the third interim analysis.

## 6.7 Sample size and operating characteristics of design options

Table 6 to Table 9 summarise the statistical performance of competing adaptive designs with two interim analyses at certain information fractions, excluding the first interim analysis at 40%. Rows marked in **red** are scenarios where the effect of DCC is the same as ECC. Rows marked on **green** are scenarios where the effect of DCC over ECC is the same as the MCID. For interpretation, let us focus on Table 6 and the rest of the tables are interpreted similarly.

Assuming an 87.2% control event rate, a maximum total of 2002 participants/mothers (1001 per arm) will be required to preserve a 90% power and a 2.5% one-sided type I error to detect a 5.2% relative increase in this event rate (4.5% absolute difference). Only around a total of 46 participants more than the sample size required for the fixed design (Section 4). This assumes two non-binding futility interim analyses at 45% (~451 per group) and 65% (651 per group) accrual of primary outcome data. The total maximum sample size is 2108 (1054 per group) accounting for a 5% expected dropout rate.

At the first interim analysis, the trial can be stopped early for futility or harm if DCC is worse than or the same as ECC; equivalent to observing a one-sided p-value of at least 0.500 or an ARD of no more than 0% (RR of  $\leq 1$ ). In addition, at the second interim analysis, the trial can be stopped early for futility if the observed effect of DCC on a critical value scale is no more

than 0.6 (~30.6% of the MCID). This is equivalent to observing no more than a 1.25% relative increase (RR of  $\leq 1.0125$ ) and at least a one-sided p-value of 0.274. Observing such a low level of evidence for DCC efficacy about two-thirds through the trial is unlikely to be overturned to show at least the targeted efficacy even if the trial progresses to the end. No early stopping for efficacy at both interim analyses is allowed. The superiority of DCC will be claimed if the critical value is above 1.96 or a one-sided p-value of less than 0.025 is observed.

If the efficacy of DCC is the same as ECC (0% RD, row in red), there is a 74.1% chance of stopping early for futility either at the first or second interim analysis. There is a 48.7% and 25.4% chance of stopping early for futility at the first and second interim analyses, respectively. The overall one-sided type I error rate is 2.4% (below the planned 2.5%). If DCC is worse than ECC by 1.2%, there is a 90.1% overall probability of early stopping; 69.0% and 21.1% at the first and second interim analyses, respectively. If the trial is stopped early at the first or second interim analysis, absolute savings in sample size will be around 901 and 700, respectively. On average, the sample size accounting for early stopping will be around 1289 and the trial would have used only 65.9% of the fixed design sample size.

On the other hand, if the targeted treatment effect is observed (row in green), there is only a negligible 2.6% chance of stopping early for futility (~1.2% at the first and ~1.3% at the second interim analyses) and a 90.3% statistical power (vs 90% planned). As the effect of DCC increases above the targeted effect (e.g., row in orange), the probability of stopping early for futility approaches 0 and power reaches 99.8%.

Table 6. Operating characteristics for interim analyses at (45%, 65%) and futility thresholds (0, 0.5) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC			Critical value	Relative risk	P-value (one-sided)		Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
85.0%	87.2%	-0.022	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	82.2 : 14.2	96.4%	0.0%	998	0.511
86.0%	87.2%	-0.012	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	69.0 : 21.1	90.1%	0.3%	1095	0.560
87.2%	87.2%	0.000	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	48.7 : 25.4	74.1%	2.4%	1289	0.659
88.0%	87.2%	0.008	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	34.4 : 23.9	58.2%	7.6%	1457	0.745
89.0%	87.2%	0.018	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	19.1 : 17.1	36.1%	23.2%	1673	0.856
90.0%	87.2%	0.028	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	8.6 : 9.1	17.7%	49.6%	1844	0.943
91.7%	87.2%	0.045	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	1.2 : 1.3	2.6%	90.3%	1980	1.013
92.5%	87.2%	0.053	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	0.3 : 0.3	0.7%	97.4%	1996	1.021
93.5%	87.2%	0.063	0.45 : 0.65	0 : 0.6	1 : 1.0125	0.5 : 0.274	1955	2002	901	1302	0.1 : 0.0	0.1%	99.8%	2002	1.024

The total maximum sample size is 2108 (1054 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.

Table 7. Operating characteristics for interim analyses at (45%, 70%) and futility thresholds (0, 0.5) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC			Critical value	Relative risk	P-value (one-sided)		Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
85.0%	87.2%	-0.022	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	82.1 : 14.5	96.7%	0.0%	1006	0.515
86.0%	87.2%	-0.012	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	68.9 : 21.6	90.5%	0.3%	1109	0.567
87.2%	87.2%	0.000	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	48.5 : 26.3	74.8%	2.4%	1305	0.667
88.0%	87.2%	0.008	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	34.4 : 24.5	58.9%	7.6%	1470	0.752
89.0%	87.2%	0.018	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	19.1 : 17.1	36.2%	23.2%	1682	0.860
90.0%	87.2%	0.028	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	8.6 : 8.5	17.1%	49.7%	1849	0.946
91.7%	87.2%	0.045	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	1.2 : 1.0	2.2%	90.3%	1974	1.010
92.5%	87.2%	0.053	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	0.4 : 0.2	0.6%	97.5%	1988	1.017
93.5%	87.2%	0.063	0.45 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1993	897	1395	0.0 : 0.0	0.1%	99.7%	1993	1.019

The total maximum sample size is 2100 (1050 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.

Table 8. Operating characteristics for interim analyses at (50%, 70%) and futility thresholds (0, 0.5) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC			Critical value	Relative risk	P-value (one-sided)		Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
85.0%	87.2%	-0.022	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	85.3 : 11.2	96.6%	0.0%	1069	0.547
86.0%	87.2%	-0.012	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	72.6 : 17.7	90.3%	0.3%	1157	0.592
87.2%	87.2%	0.000	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	51.9 : 22.5	74.5%	2.4%	1332	0.681
88.0%	87.2%	0.008	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	36.9 : 21.5	58.4%	7.6%	1487	0.761
89.0%	87.2%	0.018	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	20.4 : 15.6	36.0%	23.1%	1686	0.862
90.0%	87.2%	0.028	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	9.0 : 7.9	16.9%	49.5%	1844	0.943
91.7%	87.2%	0.045	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	1.2 : 1.0	2.1%	90.1%	1963	1.004
92.5%	87.2%	0.053	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	0.3 : 0.2	0.5%	97.5%	1976	1.011
93.5%	87.2%	0.063	0.50 : 0.70	0 : 0.6	1 : 1.0121	0.5 : 0.274	1955	1980	990	1386	0.0 : 0.0	0.1%	99.8%	1980	1.013

The total maximum sample size is 2086 (1043 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.

Table 9. Operating characteristics for interim analyses at (50%, 75%) and futility thresholds (0, 0.5) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC			Critical value	Relative risk	P-value (one-sided)		Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
85.0%	87.2%	-0.022	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	83.3 : 13.6	96.9%	0.0%	1086	0.555
86.0%	87.2%	-0.012	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	69.9 : 20.7	90.7%	0.3%	1182	0.605
87.2%	87.2%	0.000	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	48.5 : 25.7	74.2%	2.5%	1369	0.700
88.0%	87.2%	0.008	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	33.8 : 23.9	57.7%	7.6%	1523	0.779
89.0%	87.2%	0.018	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	18.1 : 16.7	34.8%	23.2%	1713	0.876
90.0%	87.2%	0.028	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	7.6 : 8.1	15.8%	49.6%	1859	0.951
91.7%	87.2%	0.045	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	0.9 : 0.8	1.8%	90.2%	1961	1.003
92.5%	87.2%	0.053	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	0.2 : 0.2	0.4%	97.5%	1971	1.008
93.5%	87.2%	0.063	0.50 : 0.75	0 : 0.6	1 : 1.0117	0.5 : 0.274	1955	1974	987	1481	0.0 : 0.0	0.0%	99.8%	1974	1.010

The total maximum sample size is 2078 (1039 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.



## 6.8 Updated sample size and operating characteristics of design options

The preliminary findings of this report (presented above) were discussed with the clinical and methodological team for feedback. Following this discussion, there was consensus within the clinical team in favour of minimising the probability of stopping early when the treatment effect is small to moderate. As such, they advised lowering the futility threshold for the second interim and preferred using a futility threshold of 0 at both the first and second interim analyses. That is, stopping the trial early for futility or harm if DCC is worse or the same as ECC at any stage of interim analyses.

Table 10 to Table 13 show the corresponding updated operating characteristics of competing adaptive designs using futility thresholds (1<sup>st</sup>, 2<sup>nd</sup>) of (0, 0) these decision rules when interim analyses (1<sup>st</sup>, 2<sup>nd</sup>) are performed at (45%, 65%), (45%, 70%), (50%, 70%), and (50%, 75%), respectively. The maximum sample sizes adjusted for a 5% expected dropout rate are presented below each table.

Table 10. Operating characteristics for interim analyses at (45%, 65%) and futility thresholds (0, 0) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC							Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
				Critical value	Relative risk	P-value (one-sided)									
85.0%	87.2%	-0.022	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	84.2 : 6.6	90.8%	0.0%	1021	0.522
86.0%	87.2%	-0.012	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	71.7 : 8.2	80.0%	0.3%	1146	0.586
87.2%	87.2%	0.000	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	52.1 : 7.8	59.8%	2.4%	1364	0.698
88.0%	87.2%	0.008	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	37.7 : 6.1	43.8%	7.6%	1532	0.784
89.0%	87.2%	0.018	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	21.9 : 3.4	25.2%	23.0%	1724	0.882
90.0%	87.2%	0.028	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	10.2 : 1.3	11.6%	49.5%	1865	0.954
91.7%	87.2%	0.045	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	1.6 : 0.1	1.7%	90.1%	1969	1.007
92.5%	87.2%	0.053	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	0.5 : 0.0	0.5%	97.4%	1981	1.013
93.5%	87.2%	0.063	0.45 : 0.65	0 : 0	1 : 1	0.5 : 0.5	1955	1986	894	1291	0.1 : 0.0	0.1%	99.7%	1986	1.016

The total maximum sample size is 2092 (1046 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.

Table 11. Operating characteristics for interim analyses at (45%, 70%) and futility thresholds (0, 0) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC			Critical value	Relative risk	P-value (one-sided)		Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
85.0%	87.2%	-0.022	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	84.2 : 8.1	92.3%	0.0%	1019	0.521
86.0%	87.2%	-0.012	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	71.8 : 10.3	82.1%	0.3%	1141	0.583
87.2%	87.2%	0.000	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	51.9 : 10.0	61.9%	2.5%	1360	0.695
88.0%	87.2%	0.008	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	37.7 : 7.8	45.4%	7.6%	1528	0.782
89.0%	87.2%	0.018	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	21.7 : 4.3	26.0%	23.2%	1723	0.881
90.0%	87.2%	0.028	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	10.2 : 1.6	11.8%	49.4%	1864	0.954
91.7%	87.2%	0.045	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	1.6 : 0.1	1.7%	90.1%	1968	1.007
92.5%	87.2%	0.053	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	0.5 : 0.0	0.5%	97.4%	1980	1.013
93.5%	87.2%	0.063	0.45 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1985	894	1390	0.1 : 0.0	0.1%	99.7%	1985	1.015

The total maximum sample size is 2092 (1046 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.

Table 12. Operating characteristics for interim analyses at (50%, 70%) and futility thresholds (0, 0) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC			Critical value	Relative risk	P-value (one-sided)		Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
85.0%	87.2%	-0.022	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	85.2 : 6.7	91.9%	0.0%	1093	0.559
86.0%	87.2%	-0.012	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	72.7 : 8.8	81.5%	0.3%	1204	0.616
87.2%	87.2%	0.000	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	51.9 : 8.8	60.7%	2.5%	1408	0.72
88.0%	87.2%	0.008	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	36.9 : 7.0	43.9%	7.6%	1567	0.801
89.0%	87.2%	0.018	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	20.5 : 3.9	24.4%	23.1%	1746	0.893
90.0%	87.2%	0.028	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	9.1 : 1.5	10.6%	49.5%	1873	0.958
91.7%	87.2%	0.045	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	1.2 : 0.1	1.3%	90.2%	1959	1.002
92.5%	87.2%	0.053	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	0.3 : 0.0	0.3%	97.4%	1968	1.007
93.5%	87.2%	0.063	0.50 : 0.70	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1380	0.0 : 0.0	0.0%	99.7%	1971	1.008

The total maximum sample size is 2076 (1038 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.

Table 13. Operating characteristics for interim analyses at (50%, 75%) and futility thresholds (0, 0) based on 500, 000 simulations.

Event rate		ARD (DCC - ECC)	Information fraction at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Futility threshold at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis:			Total SS for fixed design	Total SS with interim analyses:			Probability of futility early stopping:		Statistical power	Expected sample size	Ratio of expected SS to fixed design SS
DCC	ECC			Critical value	Relative risk	P-value (one-sided)		Maximum	1 <sup>st</sup>	2 <sup>nd</sup>	at 1 <sup>st</sup> : 2 <sup>nd</sup> interim analysis (%)	Overall			
85.0%	87.2%	-0.022	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	85.2 : 6.8	92.1%	0.0%	1098	0.562
86.0%	87.2%	-0.012	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	72.7 : 8.8	81.5%	0.3%	1212	0.62
87.2%	87.2%	0.000	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	51.9 : 8.4	60.3%	2.5%	1419	0.726
88.0%	87.2%	0.008	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	37.0 : 6.3	43.3%	7.7%	1576	0.806
89.0%	87.2%	0.018	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	20.5 : 3.2	23.7%	23.1%	1754	0.897
90.0%	87.2%	0.028	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	9.1 : 1.1	10.2%	49.4%	1877	0.96
91.7%	87.2%	0.045	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	1.2 : 0.1	1.2%	90.1%	1960	1.002
92.5%	87.2%	0.053	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	0.3 : 0.0	0.3%	97.4%	1968	1.007
93.5%	87.2%	0.063	0.50 : 0.75	0 : 0	1 : 1	0.5 : 0.5	1955	1971	986	1479	0.0 : 0.0	0.0%	99.8%	1971	1.008

The total maximum sample size is 2076 (1038 per group) accounting for a 5% expected dropout rate. Claim efficacy if critical value is above 1.96; ARD, absolute risk difference; DCC, delayed cord clamping; ECC, early cord clamping, SS, sample size; RR, relative risk/risk ratio.

## 7 Conclusions

The simulations demonstrated the impact of futility decision rules and when interim analyses are performed on the performance of an adaptive design with a non-binding futility stopping rule as well as the value of performing the second interim analysis. These results can guide the research team to select a feasible and robust adaptive design to address the research questions from a list of competing design options. Finally, this simulation report can help the research team understand aspects of the trial design and the implications of the assumptions made around the underlying treatment effect.

Of note, we assumed little uncertainty around the control event rate for the primary outcome and that sample size re-estimation is not essential as we felt that prior data that informed this rate were robust. However, the implication of this was not explored; e.g., the potential loss in power if the control rate is lower than anticipated can still be explored. Finally, we did not explore the utility of an adaptive population enrichment design due to feasibility issues as explained in the report.

## 8 References

1. Yelland, L. N. *et al.* Accounting for twin births in sample size calculations for randomised trials. *Paediatr. Perinat. Epidemiol.* **32**, 380–387 (2018).
2. Stevely, A. *et al.* An Investigation of the Shortcomings of the CONSORT 2010 Statement for the Reporting of Group Sequential Randomised Controlled Trials: A Methodological Systematic Review. *PLoS One* **10**, e0141104 (2015).
3. Wassmer, G. & Pahlke, F. RPACT Package Overview | RPACT.