SHUTEYE Trial: SHorter vs Longer CoUrse AnTibiotics – an AdaptivE BaYEsian Design Study

A platform adaptive randomised control trial in patients with Community Acquired Pneumonia (CAP) to define the optimal antibiotic duration.

Design Report

Version 2.1, October 2021

Summary of changes

Version	Changes
2.0	Updated design for stage 2 NIHR HTA application New simulations of probability of stopping at each interim analysis Allocation plan of patients to trial arms after each interim analysis added
2.1	Formatting: Page numbers added Appendix: model formulation and priors added Example trial added Formatting: figure and table numbers added

Aims and objectives

The overall objective of the trial is to define the optimal length of antibiotic treatment for patients with Community Acquired Pneumonia (CAP).

SHUTEYE is a multi-centre, randomised trial to define the minimal optimum safe antibiotic treatment length in CAP by comparing one day shorter courses against standard length using a novel design. Initially we will compare 7-days' treatment (existing standard) with 6-days of antibiotics (intervention). As a platform trial, SHUTEYE will open with two intervention arms and will add further intervention arms according to prespecified rules, using interim analyses of the accumulating data to provide statistical triggers for making these adaptations.

Population: Hospitalised Adult patients with CAP

Intervention: Reduced duration of antibiotic treatment, from 6 days to 2 days; arms opened sequentially by decreasing duration according to adaptive design rules

Comparator: Standard Length Treatment Duration of Antibiotics (starting at 7 days)

Outcome: Clinical Cure of Infection at 14 days following randomisation (binary outcome variable)

Overall design and analysis principles

SHUTEYE is a platform trial designed to identify the best duration of a course of antibiotics to treat CAP. The optimal duration will be the shortest duration that has an effect as good as the maximum effect. To estimate the duration-response curve, we will model the response across all treatment durations using a Bayesian normal dynamic linear model, which is essentially a smoothing function that does not make assumptions about the shape of the overall relationship. This will estimate the outcomes across all treatment arms and we will present the duration-response relationship and the probabilities that each duration is the best. The analysis model will also incorporate important prognostic covariates in the analysis model; these will be selected in advance and specified in the Statistical Analysis Plan, but are likely to include factors known to be related to outcomes such as age and comorbidities. We will also incorporate the effects of time; it is possible that the population or outcomes may vary through time or seasonally, and with randomisation that is not equal to all arms over time it is important to account for this in the analysis. We propose to use a similar method to that used in the PRINCIPLE trial, a recent Bayesian adaptive trial of treatments for COVID-19, which accounted for time effects by dividing the trial period into multiple sub-periods and using Bayesian hierarchical modelling. There is no final test that will establish superiority or otherwise of any particular treatment duration; instead the trial's job is to estimate the relationship, and this will form the basis of the trial's interpretation.

The trial will randomise patients to one of up to six durations of antibiotics, from 7 days to 2 days. It will take an adaptive approach to opening arms, for reasons of clinical acceptability. Clinicians will only be asked to randomise to arms with a one-day difference in duration from regimens that have evidence (from within this trial) that they are of comparable effectiveness to standard care, giving some reassurance that patients will not be randomised to a markedly inferior treatment. In addition, the difference in treatment effect between durations differing by one day is likely to be small, so patients will not be allocated to a duration of radically different duration until there is some evidence that this is not likely to be harmful.

Seven days is the currently recommended duration in national guidelines, and opening all six arms simultaneously would require clinicians to randomise patients to much shorter durations than are currently used, without any reassurance that they are likely to be as effective. Instead, the trial will take an adaptive approach to opening new treatment arms of shorter duration. Each new arm of one day shorter duration will only be opened after an interim analysis reaches a statistical trigger that shows (at a minimum) that there is not strong evidence that the shortest duration already in the trial is worse than longer durations. Demonstrating "non-inferiority" in interim analyses is unrealistic because the amount of information available will be limited, but the criterion only needs to be encouraging enough to start evaluating a new duration; subsequent interim analyses may show that it is in fact worse, at which point it may be closed.

At the same time, interim analyses may show that the longest arm currently open is worse than shorter-duration arms. The trial should allow the longest duration arm to be closed if this occurs.

Determination of operating characteristics

The design has been informed by simulation, using FACTS and R software. Because the trial design is complex and there are multiple opportunities for opening and closing arms, this design report has not yet explored all of the possible ramifications, but it provides an overall summary and the main features and expected trial operating characteristics. Additional scenarios will be added in future versions of this document.

Scenarios of treatment effect used in simulations

We studied six scenarios of treatment effect, summarised in the table below, which gives the true proportion with clinical cure for each arm in each scenario. Red figures are the duration that should be identified as the best (i.e. shortest duration that has the maximum effectiveness).

Table 1. *Scenarios of treatment effect*

Scenario name	2 day	3 day	4 day	5 day	6 day	7 day
Equal	**0.75**	0.75	0.75	0.75	0.75	0.75
Linear	0.50	0.55	0.60	0.65	0.70	**0.75**
Plateau6	0.55	0.60	0.65	0.70	**0.75**	0.75
Plateau4	0.65	0.70	**0.75**	0.75	0.75	0.75
Threshold	0.55	0.55	0.55	**0.75**	0.75	0.75
U-shaped	**0.75**	0.65	0.55	0.55	0.65	0.75

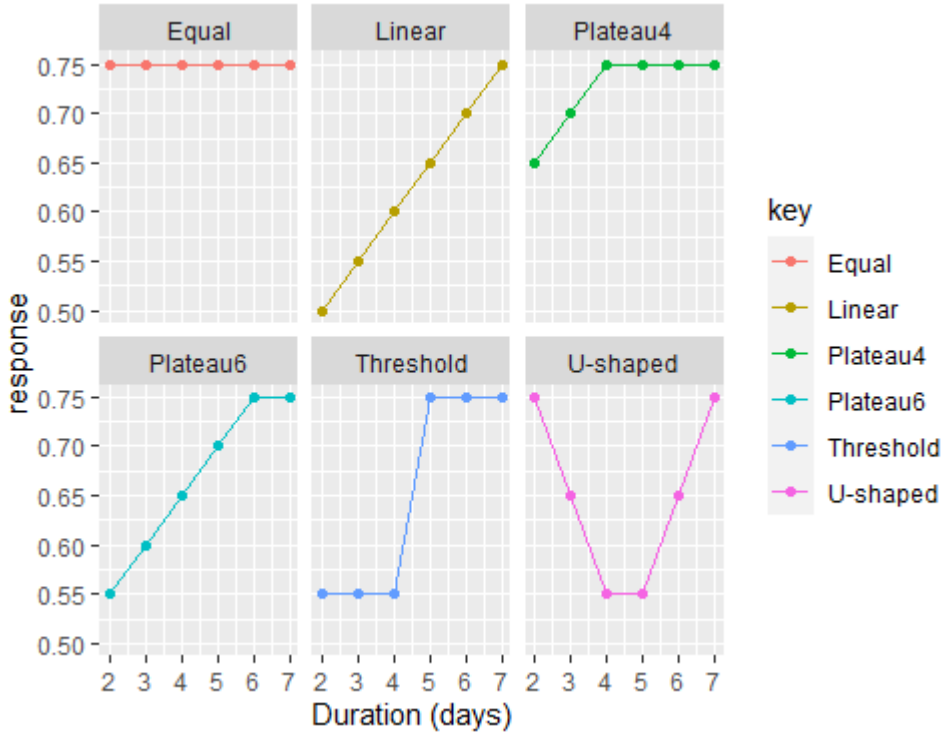

Figure 1. *Treatment effect scenarios used in simulations of trial design performance. The x-axis shows the duration of treatment in days, and the y-axis shows the proportion of patients who achieve the primary outcome, clinical cure.*

Sample size

The first set of tables below (Table 2) give the probability that each arm is the minimum duration with 95% of the maximum treatment effect (ED95), in each scenario (based on 1000 simulations), with a variety of sample sizes per arm, from 200 to 500. We assume in these simulations that each arm recruits up to the maximum sample size, in order to compare the performance of different maximum sample sizes.

Table 2. Performance of different maximum sample sizes; ED95

a. 200 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.596	0.243	0.099	0.033	0.021	0.008
Linear	0	0	0.002	0.063	0.33	0.605
Plateau6	0	0	0.021	0.206	0.623	0.15
Plateau4	0.009	0.123	0.535	0.216	0.079	0.038
Threshold	0	0	0	0.674	0.252	0.074
U-shaped	0.839	0	0	0	0.002	0.159

b. 300 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.685	0.198	0.079	0.026	0.005	0.007
Linear	0	0	0	0.027	0.346	0.627
Plateau6	0	0	0.012	0.192	0.685	0.111
Plateau4	0.003	0.119	0.634	0.177	0.051	0.016
Threshold	0	0	0	0.731	0.222	0.047
U-shaped	0.890	0.001	0	0	0	0.109

c. 400 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.753	0.169	0.05	0.018	0.006	0.004
Linear	0	0	0	0.023	0.3	0.677
Plateau6	0	0	0.004	0.189	0.719	0.088
Plateau4	0	0.094	0.706	0.152	0.04	0.008
Threshold	0	0	0	0.799	0.176	0.025
U-shaped	0.913	0	0	0	0	0.087

d. 500 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.772	0.179	0.039	0.006	0.004	0
Linear	0	0	0	0.012	0.329	0.659
Plateau6	0	0	0.002	0.159	0.772	0.067
Plateau4	0	0.066	0.767	0.138	0.027	0.002
Threshold	0	0	0	0.827	0.158	0.015
U-shaped	0.931	0	0	0	0	0.069

The following tables (Table 3) show the estimated probability of identifying each arm as having the maximum response (in other words, the estimated probability that each arm is the best), for each scenario (based on 1000 simulations).

Table 3. Performance of different maximum sample sizes: probability of best response

a. 200 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.171	0.167	0.172	0.152	0.16	0.178
Linear	0	0	0	0.004	0.092	0.904
Plateau6	0	0	0	0.028	0.455	0.517
Plateau4	0	0.007	0.228	0.241	0.253	0.271
Threshold	0	0	0	0.242	0.364	0.394
U-shaped	0.467	0.001	0	0	0	0.532

b. 300 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.172	0.169	0.18	0.156	0.151	0.172
Linear	0	0	0	0.002	0.066	0.932
Plateau6	0	0	0	0.015	0.485	0.5
Plateau4	0	0.003	0.228	0.255	0.261	0.253
Threshold	0	0	0	0.272	0.37	0.358
U-shaped	0.522	0	0	0	0	0.478

c. 400 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.166	0.166	0.177	0.161	0.158	0.172
Linear	0	0	0	0	0.052	0.948
Plateau6	0	0	0	0.009	0.449	0.542
Plateau4	0	0.003	0.228	0.248	0.253	0.268
Threshold	0	0	0	0.257	0.359	0.384
U-shaped	0.477	0	0	0	0	0.523

d. 500 per arm:

	Treatment arm					
Scenario	**2 day**	**3 day**	**4 day**	**5 day**	**6 day**	**7 day**
Equal	0.176	0.139	0.182	0.159	0.169	0.175
Linear	0	0	0	0.001	0.034	0.965
Plateau6	0	0	0	0.008	0.472	0.52
Plateau4	0	0.001	0.252	0.225	0.262	0.26
Threshold	0	0	0	0.249	0.375	0.376
U-shaped	0.489	0	0	0	0	0.511

Using the ED95 criterion, performance (the probability of selecting the correct arm as the best) improves with increasing sample size. A maximum sample size of 300 per arm appears sufficient to estimate the duration-response relationship sufficiently well to identify the best duration with high probability. 300 per arm was appreciably better than 200 per arm, but additional gains from recruiting up to 400 or 500 per arm were smaller.

The tables for the probability of each arm having the maximum response (Table 3) suggest that for this criterion, 200 per arm performs well. For example, in the "equal" scenario, all the arms have, as expected, similar probability of having the maximum response; they are all between 0.152 and 0.178. A sample size of 500 per arm does not perform better; the probabilities for the "equal" scenario range from 0.139 to 0.182 – a wider range. In other scenarios, the sample size also perform similarly; for example in the "linear" scenario, all sample sizes (correctly) have >90% probability of 7-days as having the maximum response. Similarly, all sample sizes gave approximately equal probability of having the maximum response for the correct arms in the "plateau6," "plateau4," and "threshold" scenarios (all of these had multiple duration arms with the same response).

We have therefore selected a maximum recruitment of 300 per arm, giving a total of 1800 if there is full recruitment to all 6 arms. In all scenarios this sample size selected the correct dose most often as the ED95, and there was greater than 87% probability of selecting the correct dose or a dose 1 day different.

Determination of the optimal duration at the end of the trial will not be based solely on a single statistic (such as ED95 or maximum response) but will be based on the modelled duration-treatment effect relationship. The trial's main goal, therefore, is to estimate this relationship.

Randomisation

An adaptive randomisation procedure will be used. If no arms are closed for lack of effectiveness, we expect to eventually recruit a maximum of 300 patients in each treatment allocation. Because some arms will open later than others, randomisation will be weighted towards the later-opening arms, to reach a total of 300 per arm by the end of the trial. Therefore when new arms are opened, they will initially be allocated more participants than existing arms, so that they will contain enough patients by the time of the next interim analysis to for an adaptive randomisation schedule will be used so that numbers in the newly-opened arm will have enough evaluable participants in them to be included in the following interim analysis. The numbers of patients and the ratios for randomisation at each stage is shown below:

Each period is followed by an interim analysis, with a new arm potentially opening after each of the first four. Interim analyses are therefore planned after 240, 390, 570, 780, 1020 and 1410 patients.

Table 4. Randomisation schedule

Period	7d	6d	5d	4d	3d	2d	Total per period	Cumulative total
1	120	120					240	240
2	30	30	90				150	390
3	30	30	30	90			180	570
4	30	30	30	30	90		210	780
5	30	30	30	30	30	90	240	1020
6	30	30	60	75	90	105	390	1410
7	30	30	60	75	90	105	390	1800
TOTAL per arm	300	300	300	300	300	300	1800	

If the statistical trigger for opening a new arm is not met at an interim analysis, the patients that would have been allocated to the new arm for the subsequent period will be allocated equally to all open arms. Arms will recruit up to a maximum of 300 patients.

Interim analyses and adaptation decisions

NB *There is ongoing development of the trial design. This document records the specific thresholds and design features at its specific date, but we expect that there will be further refinements and improvements, as discussions with collaborators continue and further simulations are conducted. Future changes to the design will be summarised in a new version of the design report.*

Interim analyses will be performed to take decisions about opening new duration arms and closing existing arms. At each interim analysis, a Bayesian normal dynamic linear model will be fitted across all available treatment arms, and decision-making will be based on the modelled responses. All decisions will be reviewed by the Data Monitoring Committee to check that the decisions suggested by the statistical trigger are sensible, whether the modelling and decision criteria are functioning as intended, and whether there are any other factors (including for example, data from other trials) that would influence the adaptation decisions.

Interim analyses are required fairly frequently early in the trial to allow arms to be opened rapidly, so that, later in the trial, the differential allocation is not too extreme. Our proposal (above) will give a maximum ratio of 7:2 between the most recently opened and oldest arms, which was deemed acceptable.

The trial will start with a comparison of 6-days versus 7-days. Up to six interim analyses will be conducted, when 240, 390, 570, 780, 1020, and 1410 patients have complete follow-up for the primary outcome at 14 days. The expectation is to open a new arm of 1-day shorter duration at each

of the first four interim analyses. The Data Monitoring Committee will review all statistical triggers that are met, and will make recommendations to the TSC about opening and closing of arms.

Interim analyses and adaptations

Interim analysis 1 (240 patients)

The first interim analysis will only involve 2 arms (6 days versus 7 days), and will be a simple 2-way comparison of the responses. Possible decisions at this interim analysis are:

1. Open 5-day duration arm, in addition to 6-day and 7-day arms. This will occur if we do not find convincing evidence that 7-days is superior to 6-days. The statistical trigger will be a probability of less than 90% that 7-day duration is at least 5% better (absolute) than 6-day duration.
2. Close 7-day arm; if there is strong evidence that 6-day is superior to 7-day (probability of 6-day being 5% better (absolute) than 7-day of greater than 90%).
3. Open 8-day arm; if there is a probability of greater than 90% that 7-days is at least 5% better (absolute) than 6-days. NB although in theory the trial could proceed by opening new arms of increasing duration of treatment, this is not expected and is not pursued in the remainder of this document.

Simulation results are shown only for three scenarios (Table 5); equal, linear and U-shaped, as in all others the response in the 6-day and 7-day arms are the same, and therefore the same as the equal scenario.

Table 5. Operating characteristics at interim analysis 1

	Probability of opening 5d arm	**Probability of opening 8d arm**	**Probability of closing 7d arm**
Equal	0.983	0.017	0.012
Linear	0.898	0.102	0.001
U-shaped	0.662	0.338	0

Simulation results show a high probability of opening the 5d arm if the 6-day and 7-day responses are in reality the same, but lower probabilities if in reality, 6 days is worse than 7 days. In the majority of cases the 5 day duration would open but there are opportunities for it to be stopped at subsequent interim analyses, when more information would probably show it to be inferior.

Interim analyses 2, 3 and 4

For the second to fourth interim analyses, if three or more arms are open, the Bayesian normal dynamic linear model will be estimated across all durations and used for decision making. At each interim the possible decisions are:

1. If the shortest duration currently open is not worse than longer durations, a new arm with 1 day shorter duration will be opened, with a minimum duration of 2 days. The criterion used will be a probability greater than 0.2 that the shortest duration is the ED90, the shortest duration which achieves at least 90% of the maximum treatment effect.
2. If the longest duration arm currently open has a low probability of having the best estimated response (probability that it is the arm with maximum response < 0.01), that arm may be closed.

3. If the shortest duration arm currently open has a low probability of having the best estimated response (probability that it is the arm with maximum response < 0.01), that arm may be closed.
4. If there is high probability that the longest duration currently open is the best (probability of having the maximum treatment effect > 0.99), the trial may terminate.

If no decision criteria are reached, the trial will continue to randomise to the existing arms until the next interim analysis, with patients that would have been allocated to the new arm being divided equally between the open arms.

Table 6 presents the probabilities of the possible decisions at each interim analysis (based on 1000 simulations). All of these figures are based on the assumption that at earlier interim analyses (a) a new arm of 1 day shorter duration is opened; (b) no arms have been closed at earlier interim analyses; (c) randomisation is according to the table given above.

Table 6. Operating characteristics for interim analyses 2, 3 and 4

	Probability of opening new arm (1 day shorter)	**Probability of closing shortest arm**	**Probability of closing 7d arm**	**Probability of terminating trial**
Interim 2				
Equal	0.980	0.026	0.031	0
Linear	0.666	0.310	0	0.034
Plateau6	0.877	0.149	0.012	0.001
Plateau4	0.980	0.026	0.031	0
Threshold	0.980	0.026	0.031	0
U-shaped	0.124	0.828	0	0.284
Interim 3				
Equal	0.981	0.028	0.056	0
Linear	0.289	0.695	0	0.043
Plateau6	0.525	0.512	0.005	0.001
Plateau4	0.981	0.028	0.056	0
Threshold	0.035	0.968	0.02	0.001
U-shaped	0.093	0.873	0	0.314
Interim 4				
Equal	0.966	0.064	0.103	0
Linear	0.056	0.941	0	0.066
Plateau6	0.179	0.844	0.007	0
Plateau4	0.792	0.283	0.063	0
Threshold	0.017	0.987	0.022	0
U-shaped	0.603	0.356	0	0.187

These simulation results suggest that the design will perform well. There is a high probability of opening a new arm (1 day shorter duration) when that is appropriate. For example, in the "equal" scenario, it should always be appropriate to open a new arm (as all durations have the same effectiveness), and this probability is high across all interim analyses. In the "Linear" scenario, 7-day duration is the best, and here there is a much lower probability (0.666) of opening a 5-day arm at

interim 2, and this probability gets lower with subsequent interims. The probability that the 5-day arm is subsequently closed in this scenario is high (0.695 at interim 3).

Probabilities of closing the longest (7-day) arm are low, and are zero in scenarios where 7 days is in fact the most effective duration.

The probabilities of closing the shortest arm when that is in fact as good as the best duration are low (0.02 to 0.03) for (equal, plateau4 and threshold for interim 2, and equal and plateau4 for interim 3). Where the shortest duration is in fact worse than other arms in the trial, the probability of closing is much higher e.g. 0.695 for the 5-day arm in the linear scenario at interim 3, and 0.844 for the 3-day arm at interim 4 in the plateau6 scenario, where 6 days is in fact the optimal duration.

The results for the U-shaped scenario are interesting, and suggest that the design may fail if there is this sort of relationship between duration and response. Here, 2 days and 7 days are the best durations, with all others being inferior. In this scenario there is a low probability of opening new arms at the first few interim analyses, and high probabilities of closing the shortest arm, because 6, 5, and 4 days are all inferior to 7 days. It is likely that the shortest duration arms (2 and 3 days) would never be opened , meaning there would be a high probability of failing to identify that 2 days was as good as 7 days. However, this scenario is biologically implausible, and it is much more likely that the response increases with increasing duration; the question is about the shape of that relationship.

Interim analysis 5 (1020 patients)

The decision criteria to be applied at the fifth interim analysis will depend on how many arms are open. If fewer than 5 arms are open, the same criteria will be applied as at interims 2, 3 and 4. The fifth interim analysis will be the last occasion that new arms may be opened. If all six arms are open, the decision criteria available at this interim analysis will be:

1. If there is high probability that the longest duration currently open is the best (probability of having the maximum treatment effect > 0.99), the trial may terminate.
2. If the shortest duration arm currently open has a low probability of having the best estimated response (probability that it is the arm with maximum response < 0.01), that arm may be closed.

If all six arms are not open at this point the trial will run to completion (300 per arm) with fewer than six arms.

Interim analysis 6 (1410 patients)

The final interim analysis (1410 patients) may not be conducted if there are not six open arms, and hence it is close to the end of recruitment (1500 patients with 5 arms). Decision criteria available at this interim analysis will be:

1. If there is high probability that the longest duration currently open is the best (probability of having the maximum treatment effect > 0.99), the trial may terminate.
2. If the shortest duration arm currently open has a low probability of having the best estimated response (probability that it is the arm with maximum response < 0.01), that arm may be closed.

Table 7 gives the probabilities of these actions for each scenario, for interims 5 and 6 (combined):

Table 7. Operating characteristics for interim analyses 5 and 6.

	Probability of closing shortest arm	**Probability of terminating trial**
Interims 5 and 6		
Equal	0.127	0
Linear	0.784	0.101
Plateau6	0.844	0.006
Plateau4	0.876	0.001
Threshold	0.743	0.002
U-shaped	0.001	0.006

Analysis strategy

The final analysis will use a Bayesian normal dynamic linear model to estimate the outcomes across all treatment arms and will present the duration-response relationship and the probabilities that each duration is the best. We will also incorporate important prognostic covariates in the analysis model; these will be selected in advance and specified in the Statistical Analysis Plan, but are likely to include factors known to be related to outcomes such as age and comorbidities. We will also incorporate the effects of time; it is possible that the population or outcomes may vary through time or seasonally, and with randomisation that is not equal to all arms over time it is important to account for this in the analysis. We propose to use a similar method to that used in the PRINCIPLE trial,[1] a recent Bayesian adaptive trial of treatments for COVID-19, which accounted for time effects by dividing the trial period into multiple sub-periods and using Bayesian hierarchical modelling.

An important aspect of using an NDLM is to employ priors for the relevant parameters that will give an appropriate degree of smoothing over the duration arms. Different values can either accentuate or reduce the amount of smoothing. More smoothing is helpful in the null scenario, as it removes random fluctuations when all arms are actually the same, but is less helpful when there are large jumps between adjacent durations; this can be seen in the simulation results in the Sample Size section above for the "Threshold" scenario, in which there is a large jump in response between 4 days and 5 days. However, 6 days rather than 5 days was identified as the ED95 duration in 15-25% of simulations, and 5 days was identified as the duration with best response in fewer simulations than 6 days. These effects are caused by smoothing, causing the estimated effect for 5 days to be lower than that of 6 days. Priors for smoothing parameters will be investigated in a separate simulation exercise, which will be reported in the Statistical Analysis Plan.

Illustrative simulated trial examples

In this section we provide examples of simulated trials, showing how the accumulating data are used to make decisions about opening new arms at interim analyses.

Example 1

The first example trial is from the "linear" scenario, where effectiveness increases linearly with increasing duration of antibiotic course. In reality the 7-day course is the best, with effectiveness declining with every day shorter.

Interim analysis 1

Initially, randomisation is to two arms, 6 days and 7 days, with the first interim analysis carried out when the primary outcome for 240 patients is known. The results at this analysis are presented in Table 8.

Table 8. Results of interim analysis 1

	6 day course	**7 day course**
Proportion with outcome (95% CI)	0.699 (0.616 – 0.774)	0.716 (0.629 – 0.798)
True (unknown) proportion	0.70	0.75
Probability of having the best response	0.389	0.611
Probability of being at least 5% better than the other arm	0.1278	0.2824

Applying the decision criteria at this interim analysis:

1. There is not clear evidence that the 7-day arm is better; the probability that it is at least 5% better is 0.2824, which is less than the threshold of 0.90. Hence the 5-day arm would be opened after this interim analysis.
2. There is not clear evidence that the 6-day arm is superior to 7 days (probability of being at least 5% better is 0.1278), hence the 7-day arm will remain open until (at least) the next interim analysis.
3. Because there is not strong evidence that the 7-day arm is superior, there is no need to consider opening an 8-day arm.

Interim analysis 2

Recruitment continues to 3 arms, with patients allocated to 5, 6, and 7 days in the ratio 3:1:1, giving totals of 90, 150 and 150 in the three arms at the next interim analysis. Results of this interim analysis are shown in Figure 2 and Table 9.

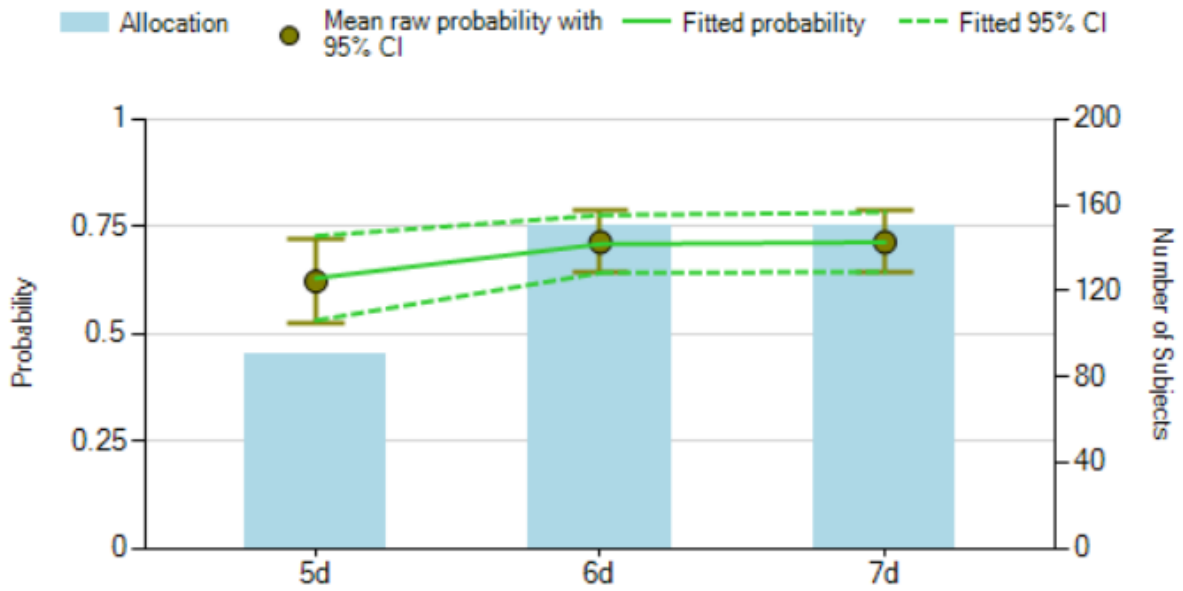

Figure 2. *Results of interim analysis 2, showing number of patients allocated to each arm (blue bars), response and 95% CI in each arm (dark green points and bars) and modelled estimates and 95% CI (bright green solid and dotted lines).*

Table 9. Results and statistical triggers for interim analysis 2

	5 day course	**6 day course**	**7 day course**
Number of patients	90	150	150
Proportion with outcome (95% CI)	0.63 (0.53 – 0.72)	0.71 (0.64 – 0.77)	0.71 (0.64 – 0.78)
True (unknown) proportion	0.65	0.70	0.75
Probability of having the best response	0.04	0.45	0.52
Probability of being shortest duration with 90% of maximum response (ED90)	0.32	0.62	0.06

1. The 7-day duration does not have a sufficiently high probability of being the best arm (probability is 0.52) to terminate the whole trial, so randomisation will continue.
2. Similarly, the 7-day arm has a probability of greater than 0.01 of being the best arm, so it will not be closed.
3. 5 days has a probability of being the ED90 duration of 0.32, which exceeds the threshold of 0.2, so a further arm with 4-day duration will be opened.
4. The probability that the shortest arm (5 days) is the best is low (0.04) but the threshold of 0.01 is not reached, therefore this arm will remain open.

Interim analysis 3

Randomisation proceeds with four arms open, with an allocation ratio of 3:1:1:1 to the 4, 5, 6, and 7-day arms respectively. Interim analysis 3 is conducted when 570 patients have outcome data (Figure X). With the larger number of patients included in this analysis, the modelled duration-response curve is getting closer to the true relationship (Figure 3).

Figure 3. *Results of interim analysis 3. See above for key.*

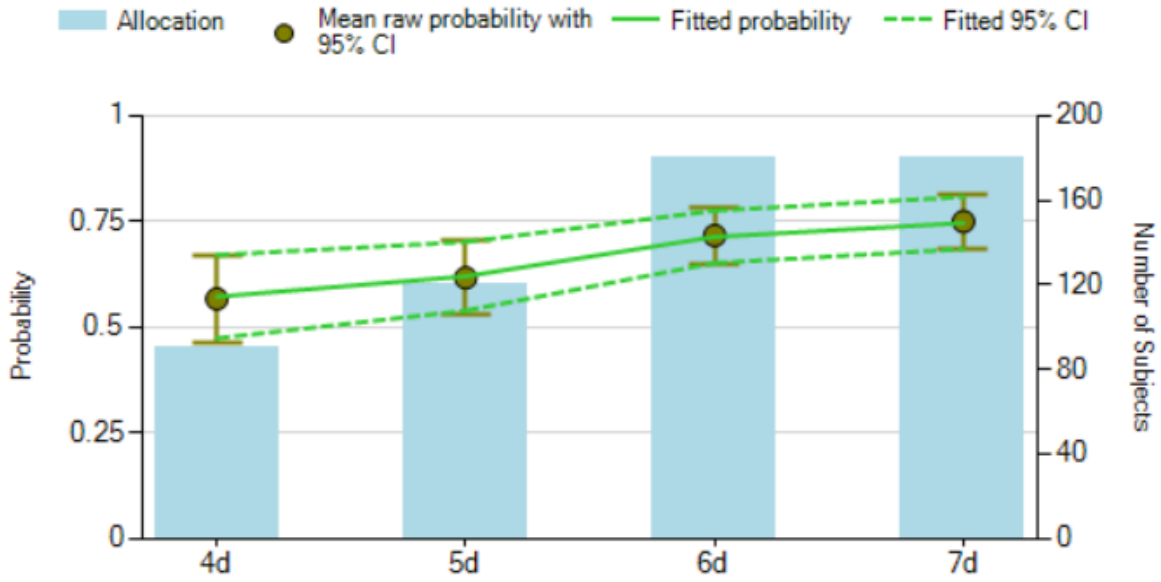

Table 10. Results and statistical triggers for interim analysis 3

	4 day course	5 day course	6 day course	7 day course
Number of patients	90	120	180	180
Proportion with outcome (95% CI)	0.57 (0.47 – 0.67)	0.62 (0.53 – 0.70)	0.71 (0.65 – 0.77)	0.75 (0.68 – 0.81)
True (unknown) proportion	0.60	0.65	0.70	0.75
Probability of having the best response	0.0002	0.01	0.22	0.77
Probability of being shortest duration with 90% of maximum response (ED90)	0.03	0.11	0.72	0.15

1. The probability of being the best arm for the 7-day duration is 0.77, which is not high enough to terminate the trial, so randomisation will continue, and the 7-day arm will not be closed.
2. The 4-day arm has a probability of being the ED90 duration of 0.03, which is below the threshold of 0.2, so a new 3-day arm will not be opened.
3. The probability that the 4 day arm is the best is very low (0.0002) and below the threshold of 0.01, so this arm will be closed.

The trial will therefore continue with 3 arms open: 5, 6 and 7 days.

Interim analysis 4

Because the 4-day arm was closed, randomisation continues with 3 arms. This necessitates a departure from the randomisation schedule, which assumed that arms would always be opened

where possible. In this scenario, we know that 7 days is the best and shorter arms are progressively worse, so there is a strong possibility that shorter arms will not open. This is appropriate because they are (in reality) inferior, and avoiding opening them ensures better treatment for patients in the trial. There are 210 patients to be allocated before the next interim analysis: allocating them 90:60:60 to the 5, 6 and 7-day arms will start to equalise the numbers and give 210, 240 and 240 in the three arms for interim analysis 4.

Although no more patients are randomised to the 4-day arm, it is retained in the analysis. Results of interim analysis 4 are given in Figure 4 and Table 11.

Figure 4. *Results of interim analysis 4*

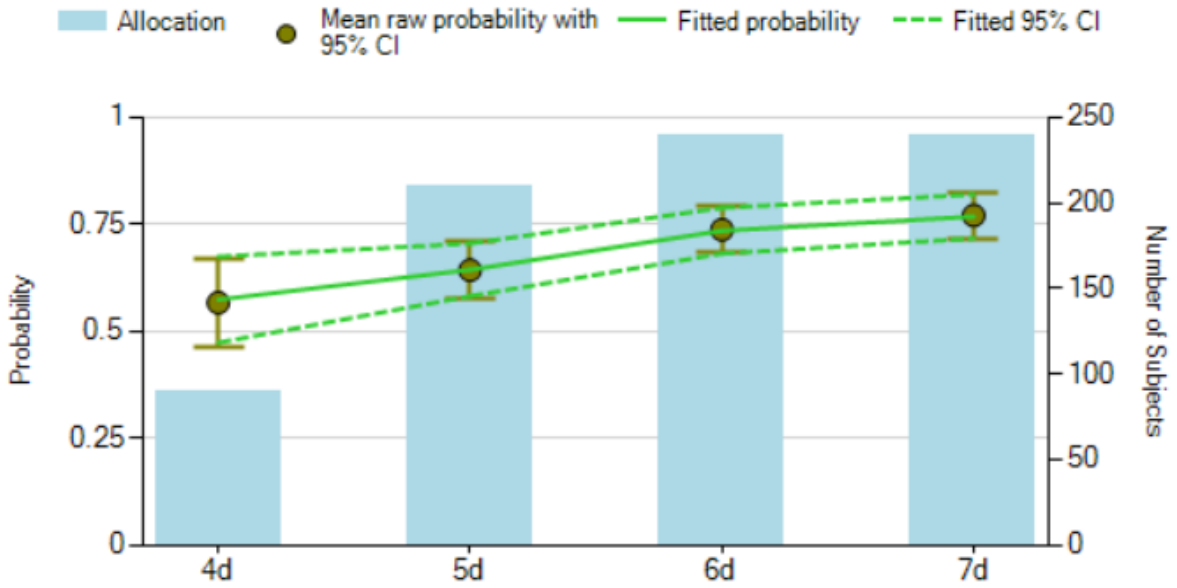

Table 11. *Results of interim analysis 4*

	4 day course	5 day course	6 day course	7 day course
Number of patients	90	210	240	240
Proportion with outcome (95% CI)	0.57 (0.47 – 0.67)	0.64 (0.58 – 0.70)	0.74 (0.68 – 0.79)	0.77 (0.72 – 0.82)
True (unknown) proportion	0.60	0.65	0.70	0.75
Probability of having the best response	0	0.0002	0.1754	0.8244
Probability of being shortest duration with 90% of maximum response (ED90)	0.0126	0.082	0.8078	0.0976

The 5-day arm has now also reached the statistical trigger of a probability of less than 0.01 of being the best arm, so is closed at this point. Randomisation can continue with just two arms, 6 days and 7 days. The 7-day arm has not reached any triggers for closure or termination of the trial.

Both of the remaining arms reach the maximum per arm of 300 when 60 patients have been randomised to each. The trial terminates at this point, and the final analysis is conducted.

Final analysis

The final analysis contains all 4 arms that have been opened, with between 90 and 300 patients in each, and 900 patients in total.

Figure 5. *Results of final analysis*

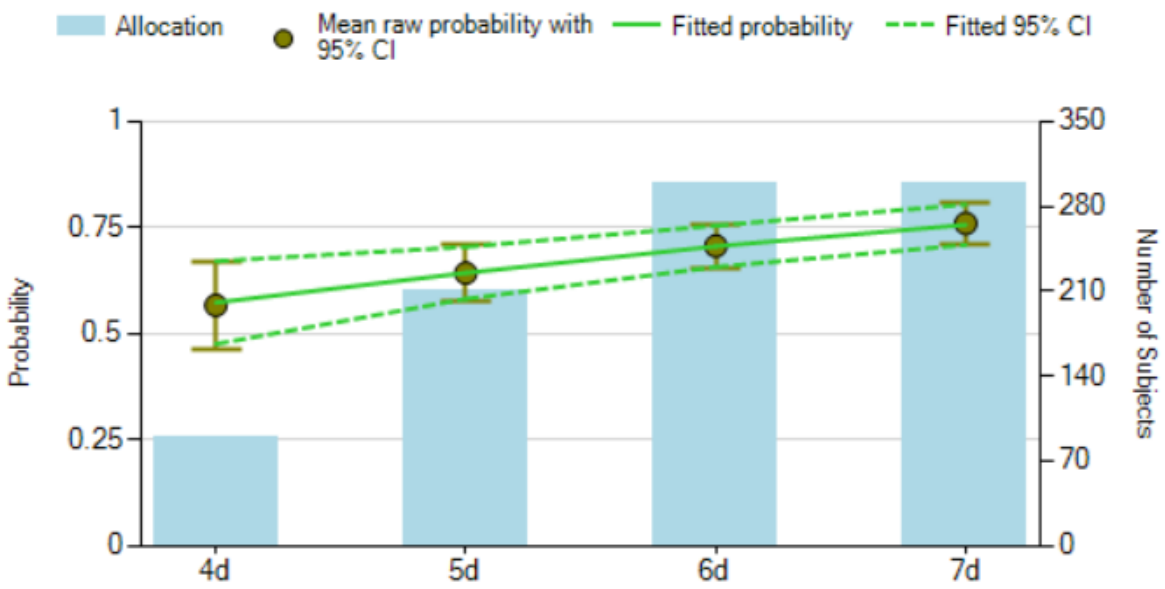

Table 12. *Results of final analysis*

	4 day course	**5 day course**	**6 day course**	**7 day course**
Number of patients	90	210	300	300
Proportion with outcome (95% CI)	0.57 (0.47 – 0.67)	0.64 (0.58 – 0.70)	0.71 (0.66 – 0.75)	0.75 (0.71 – 0.81)
True (unknown) proportion	0.60	0.65	0.70	0.75
Probability of having the best response	0.0002	0.0002	0.06	0.94
Probability of being shortest duration with 90% of maximum response (ED90)	0.0216	0.147	0.6266	0.2048

The final analysis shows that the 7-day course has by far the highest probability of being the best intervention (0.94). This conclusion was reached with half of the planned maximum number of patients (900 rather than 1800).

References

Hansen CH et al. Development of a Bayesian response-adaptive trial design for the Dexamethasone for Excessive Menstruation study. Stat Methods Med Res. 2017 Dec; 26(6): 2681–2699.

Yu L et al. Inhaled budesonide for COVID-19 in people at high risk of complications in the community in the UK (PRINCIPLE): a randomised, controlled, open-label, adaptive platform trial. Lancet 2021; 398: 843–55

APPENDIX: Model formulation and priors

The normal dynamic linear model (NDLM) is a Bayesian smoothing function, that fits a smooth curve to the data across all of the treatment durations. The model makes no assumptions about the shape of the duration-response relationship, so is appropriate when we wish to allow for the curve to be any shape, including having a peak an intermediate value, or declining effectiveness with increasing duration. Such shapes are considered unlikely but we want to allow for the possibility that they could occur. As with other Bayesian modelling techniques, the NDLM borrows information from neighbouring durations to estimate the response of a particular treatment arm.

For the primary outcome, which is a proportion, the NDLM relates the log odds of the outcome to the duration arms. In the model, the shortest duration has a normal prior for the log odds of the response:

$$\theta_d \sim (\mu_{0d}, \nu_{0d}^2)$$

θ_d is the response in duration 1 arm
μ_{0d} is the prior mean
ν_{0d} is the prior standard deviation

In simulations we used a mean of 0 and a standard deviation of 2 for the prior for the shortest duration. This is a standard non-informative prior. The priors for other durations are based on the response of the preceding duration. Again, they are normal priors for the log odds of the response at that duration:

$$\theta_d \sim (\theta_{d-1}, \tau_d^2)$$

θ_{d-1} is the response for the duration 1 day shorter
τ_d^2 is the between-durations variance. This is the same for all durations, because the steps between the durations are all of equal size, and has an inverse gamma prior. This parameter affects the amount of smoothing that is applied from duration to duration; a small value will "iron out" variations between durations, whereas a large value will allow apply little or no smoothing. As with other parameters in Bayesian modelling, the between-duration variance in response is updated by the observed data, but because there are a small number of duration arms, there is a limited amount of information about this parameter, so the choice of prior is likely to have some residual effect. In the simulations we used a moderate value; an inverse gamma prior with a mean of 1 and weight of 1.