eAppendix 1 – Brain atrophy modelling

Rate of whole brain atrophy reported in previous trials

***Table 1**. Rates of whole brain atrophy in progressive MS from trials using directly measured change*

					Atrophy rate (%/year)			
					Placebo		Active	
Phase	Trial	Population	MRI time points	Duration (years)	Mean	SD	Mean	SD
II	MS-STAT1	SPMS	Baseline, 12 months	1.1	-0.60	0.62	-0.38	0.63
			Baseline, 25 months	2.1	-0.56	0.47	-0.29	0.51
II	Lamotrigine	SPMS	Baseline, 12 months	1	-0.54	0.67	-0.82	0.71
			Baseline, 24 months	2	-0.59	0.37	-0.67	0.39
II	MS-SMART	SPMS	Baseline, week 96	1.85	-0.70	0.60		
II	Lipoic acid	SPMS	Baseline, 24 months	2	-0.65		-0.21	
III	CUPID**	PPMS	Baseline, 12 months	1	-0.59	0.95	-0.60	0.99
			Baseline, 24 months	2	-0.65	0.95	-0.58	0.96
			Baseline, 36 months	3	-0.76	1.04	-0.88	0.87
III	INFORMS	PPMS	Baseline, 36 months	3	-0.51		-0.50	
			Baseline, 12 months*	1	-0.55	0.67		
			Baseline, 24 months*	2	-0.52	0.51		
			Baseline, 36 months*	3	-0.50	0.40		
III	ORATORIO	PPMS	week 24, week 120	1.85	-0.59		-0.49	
			Baseline, week 24	0.50	-0.79	1.10	-0.81	1.16
			Baseline, week 48	0.97	-0.64	0.78	-0.66	0.76
			Baseline, week 120	2.36	-0.58	0.46	-0.55	0.46
III	ASCEND	SPMS	Week 24, week 96	1.38	-0.52	0.48	-0.48	0.43
			Baseline, week 24	0.50	-0.53	0.93	-0.73	1.01
			Baseline, week 48	0.97	-0.50	0.59	-0.58	0.57
			Baseline, week 72	1.38	-0.48	0.49	-0.56	0.45
			Baseline, week 96	1.85	-0.48	0.41	-0.51	0.39
III	EXPAND	SPMS	Baseline, 12 months	1	-0.46		-0.28	
			Baseline, 24 months	2	-0.42		-0.36	

*Data for Placebo and active arms combined

** CUPID was the only trial where the SD did not appear to decrease with duration of follow-up.

Expected atrophy rate variance for different follow-up lengths

Statistical model

Due to the lack of data on intermediate follow-up length between 1 and 2 years, further modelling was conducted to estimate how the standard deviation would be expected to vary over this window. This approach uses the following model for repeated measures of brain atrophy[1]:

$z_{ijk} = (\beta_1 + b_i)(t_{ik} - t_{ij}) + p_{ij} - p_{ik} + q_{ijk}$

z_{ijk} - atrophy (%) between the *j*th and *k*th visit for the *i*th participant

$t_{ik} - t_{ij}$ - time between *j*th and *k*th MRI scan

Random slope: $b_i \sim N(0, \sigma_b^2)$

Random effect for visit specific deviations from linearity: $p_{ij},\ p_{ik} \sim N(0, \sigma_p^2)$

[1] Frost C, Kenward MG, Fox NC. The analysis of repeated 'direct' measures of change illustrated with an application in longitudinal imaging. Stat Med. 2004 Nov 15;23(21):3275-86.

Residual error in measuring the change: $q_{ijk} \sim N(0, \sigma_q^2)$

In order to obtain estimates of each of the variance components the model was fitted using data from previous clinical trials. One set of estimates were obtained from the MS-STAT1 study (N=140), which had atrophy measured between each pair of scans (baseline and 12 months, 12 and 25 months, baseline and 25 months).

A second set of estimates were obtained from the ASCEND clinical trial (N=814), which had repeated measures of atrophy between baseline and each follow-up visit (baseline to 24, 48, 72 and 96 weeks). As all measures were changes from baseline the mixed model for directly measured change was simplified to:

$$z_{ik} = (\beta_1 + b_i)(t_{ik}) + p_i + \varepsilon_{ik}$$

The random slope is as defined previously with:
t_{ik} - time between baseline and follow-up MRI at the *k*th visit
p_i – random effect for baseline deviation from linearity
ε_{ik}- residual error due deviations from linearity at *k*th visit and error due to measuring the change

From the MS-STAT1 data, where the measures were between pairs of scans, the predicted variance of a single measure of atrophy rate (%/year) between baseline and follow-up at time t is:
$\sigma_t^2 = \sigma_b^2 + \frac{2\sigma_p^2 + \sigma_q^2}{t^2}$

Using the ASCEND data, where the measures were all from baseline, the predicted variance of a single measure of atrophy rate (%/year) between baseline and follow-up at time t is:
$\sigma_t^2 = \sigma_b^2 + \frac{\sigma_p^2 + \sigma_\varepsilon^2}{t^2}$

Results

The predicted standard deviation based on the data from MS-STAT1 and ASCEND is shown in Figure 1. Predictions from the model in both trials closely matched the observed standard deviations. For example, in MS-STAT1 the predicted standard deviation was 0.67 %/year for 1.1 years follow-up, which compares to the observed values of 0.62 and 0.63 %/year in placebo and active arms respectively, and the predicted standard deviation was 0.48 %/year for 2.1 years follow-up, which compares to observed values of 0.47 and 0.51 %/year for placebo and active arms respectively.

In general, the predicted standard deviation was higher based on MS-STAT1 than based on ASCEND. From MS-STAT1 it would be expected that the standard deviation would be 0.72 %/year for 1 year follow-up, 0.56 %/year for 1.5 years and 0.49 %/year for 2 years. From ASCEND the predictions are 0.59%/year, 0.46 %/year and 0.41 %/year for 1, 1.5 and 2 years follow-up length respectively.

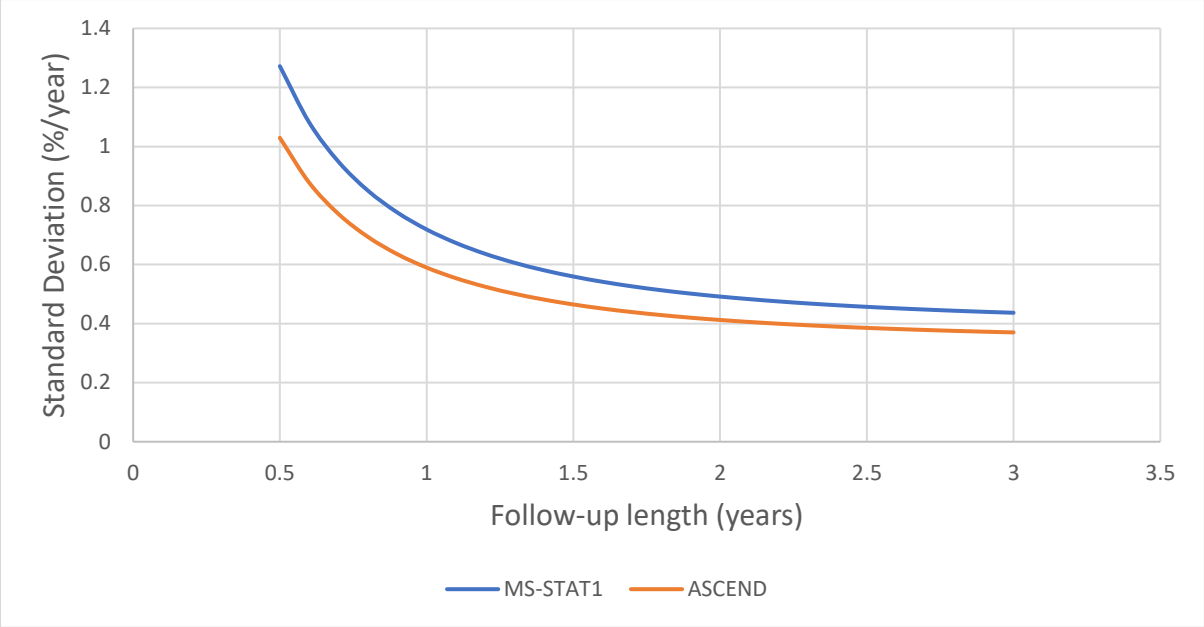


Figure 1. *Predicted standard deviation of atrophy rate for varying follow-up length, based on modelling of MS-STAT1 and ASCEND*

eAppendix 2 – Association between treatment effect on brain atrophy and on disability progression

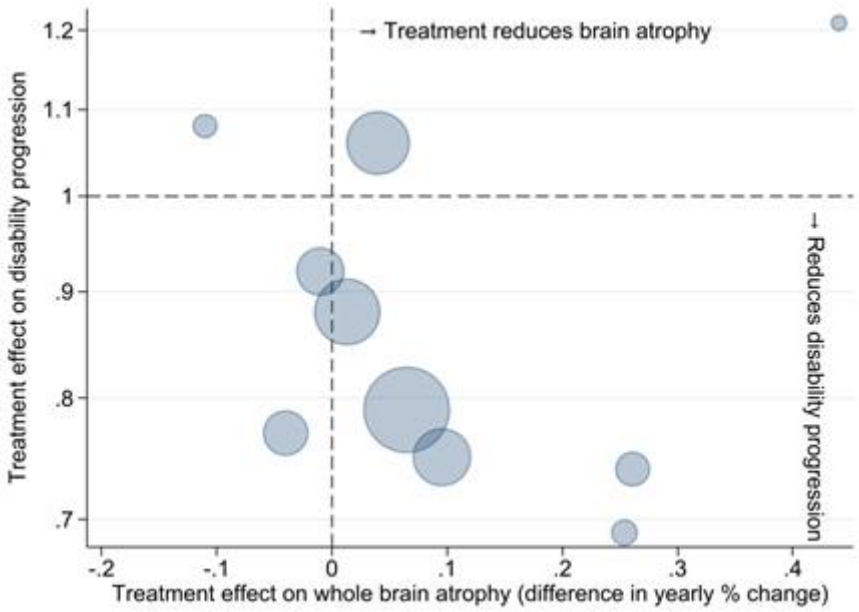
Information on progression rates and brain atrophy was extracted from the studies included in the review of randomised controlled trials in PMS. Trials which reported treatment effect on both brain atrophy and disability progression were included, to assess the association between the proposed interim and final outcomes. The specific atrophy and progression measures reported differed between studies and information was extracted which aligned most closely to the following definitions:
Brain atrophy: Mean difference in yearly percentage change in whole brain volume from baseline.
Disability progression: Hazard ratio for time to 6-months confirmed disability progression on EDSS. EDSS-based progression was used as reference as most commonly used measure across studies.

Table 1. *Treatment effect on brain atrophy and clinical progression in previous trials.*

Trial	**Whole brain atrophy (% change/year)**		**Disability progression (based on EDSS)**	
	Mean difference [95% CI]	**Derivation (from published figures)**	**Effect (HR) [95% CI]**	**Derivation (from published figures)**
Phase-III				
INFORMS	0.013	Based on % change at 3 years	0.88 [0.72 to 1.08]	HR for 3-months confirmed EDSS progression
EXPAND	0.065	Based on % change at 2 years	0.79 [0.65 to 0.95]	HR for 3-months confirmed EDSS progression
ORATORIO	0.10	Based on % change at 2 years	0.75 [0.58 to 0.98]	HR for 6-months confirmed EDSS progression
ASCEND	0.04	Based on % change between 6 and 24 months	1.06 [0.74 to 1.53]	Odds ratio for 6-months confirmed EDSS progression at 2 years
CUPID	-0.01 [-0.26 to 0.24]	Yearly % change (over 3 years)	0.92 [0.68 to 1.23]	HR for 6-months confirmed EDSS progression
OLYMPUS	-0.040	Approximation, using mean change divided by baseline volume	0.77 [0.55 to 1.09]	HR for 3-months confirmed EDSS progression
Phase-II				
SPRINT-MS	0.26	Approximation, based on change in brain parenchymal fraction	0.74 [0.43 to 1.28]	HR for 5-months confirmed EDSS progression
MS-STAT1	0.254 [0.09 to 0.42]	Yearly % change (over 2 years)	0.69 [0.42 to 1.13]	HR for unconfirmed EDSS progression. Based on data re-analysis, not published.
Lamotrigine	-0.11 [-0.26 to 0.04]	Yearly % change (over 2 years)	1.08 [0.59 to 1.99]	HR for unconfirmed EDSS progression. Based on data re-analysis, not published.
Lipoic Acid	0.44 [0.16 to 0.73]	Yearly % change (over 2 years)	1.21	Risk ratio for unconfirmed EDSS deterioration

HR= hazard ratio, 95% CI = 95% confidence interval.

Figure 1. *Association between treatment effect on brain atrophy and disability progression.*



The size of each circle is proportional to the trial size (number of patients randomised). See above for the definition of treatment effect.

Conclusion

The results of previous trials indicate an association between effect on brain atrophy and disability progression. The treatments which demonstrated a larger effect on brain atrophy tended to have a larger effect on disability progression.

eAppendix 3 – Rate of confirmed disability progression in previous trials

Methods

Information on progression rates was extracted from the studies included in the review of randomised controlled trials in PMS. Data were extracted on progression rates from trials which had a follow-up of at least 3 years, and reported the probability of clinical progression at 3 years in a table or as a figure (e.g. Kaplan-Meier plot). This was used to compare progression rates between different outcome definitions and to estimate proportion of participants expected to experience confirmed disease progression in the trial.

Definitions of disability progression differed between trials, and was either based on EDSS scale, or on a composite measure, defined as a progression on either EDSS, 9HPT or T25FW scores. The progression event was confirmed at a subsequent visit, usually three or six months later. The approximate probability of progression at three years was read from the Kaplan Meier curves or from the tables if reported. When the progression rate differed between arms, estimates were based on the active treatment arm as these are more likely to be closer to contemporary progression rates.

Results

Table 1. Probability of confirmed clinical progression within three years in previous trials in progressive multiple sclerosis.

	EDSS progression		**Composite progression****	
Trial	**3M-confirmed**	**6M-confirmed**	**3M-confirmed**	**6M-confirmed**
ASCEND	-	-	-	52%
EXPAND	35%	25-30%	-	-
INFORMS	50%	40%*	75%	60%*
ORATORIO	35%	30%	60%	55%
CUPID	-	45%	-	-

* 6-months confirmation data not reported, extrapolating from 3-months confirmation data (assuming 20% less confirmed at 6-months, based on trials were both were reported)
** Confirmed progression on either EDSS, 9HPT or T25FW

More events are expected for the composite-based progression as opposed to the EDSS-only progression, and 6-months confirmation results in slightly lower rates than 3-months confirmation (Table 1). The progression rate for the primary outcome of composite progression with 6 months' confirmation is between 52-60% at 3 years. On this basis, a conservative (lower) progression rate estimate of 50% at 3 years has been used for sample size calculations, which allows for the broader EDSS inclusion criteria in this trial compared to previous studies, and that participants may receive a newly licensed treatment in addition to the active medication being investigated. Both of these factors may lead to slightly lower progression rates than observed previously.

eAppendix 4 – Trial timeline

Methods

Predicted trial progress was modelled in Microsoft Excel, modelling for each month relevant indicators such as the expected number of participants recruited in the trial, or the expected number of participants experiencing the primary outcome.

Follow-up design

Variable and fixed duration of follow-up were considered. Under fixed duration, all participants would be followed up for the same length of time (e.g. 4 years from recruitment) and the stage 2 analysis performed once the last patient reached their final visit. With variable duration, participants are followed up until the required number of primary outcome events is reached (Figure 1).

A variable follow-up design was selected for the trial, as the required number of events for the final analysis is reached in a shorter timescale. However, variable follow-up could lead to very long follow-up length for patients recruited early in the trial. Based on feedback from the PPI group, it was decided to set a maximum follow-up length of 5 years, as longer durations may be a barrier to trial recruitment.

Aside of trial drop-out, participants are therefore anticipated to remain in the trial until i) their arm is stopped at stage 1 analysis, or ii) they complete 5 years of follow-up, or iii) the required number of progression events is observed and stage 2 analysis conducted (whichever occurs first).

Figure 1: Possible follow-up schedules. A = fixed duration, B = variable duration with maximum duration

Parameters

We modelled the progress of the trial and the expected overall duration under a 'base-case' scenario, and under alternative scenarios. The parameters assumed in the model are summarised in Table 1

Table 1. *Summary of parameters assumed in the trial timeline model.*

Parameter	**Base-case value**	**Alternative value(s)**	**Notes**
Number of experimental arms at start of trial	3 arms		
Allocation ratio	1:1:1:1		Equal allocation between each trial arm
Recruitment rate (participants per month)	40 during stage 1 and 50 in stage 2	● 20% lower (32,40) ● 20% higher (48,60)	Assuming recruitment starting slower in first year, increasing proportionally from 0 to full recruitment rate in 12 months. Also assuming recruitment to be faster during stage 2.
Drop-out	10%		Assuming 10% of recruited participants drop-out before 18M MRI data or confirmed clinical progression
Number of participants per arm	600		Stop recruitment when reached 600 participants per arm
Maximum follow-up duration	5 years		Participant exit the trial after a maximum of 5 years of follow-up
Interim analysis (stage 1)			
Interim outcome time-point	18 months		Interim analysis conducted when 18-months MRI data are available for 111 patients per arm
Interim analysis sample size	111 per arm		
Delay between data collection and analysis	4 months		Time to process the MRI data, conduct the analysis, and effectively stop the arms.
Final analysis (stage 2)			
Number of experimental arms in second stage	1 arm	● 2 arms ● 3 arms	
Clinical progression rate in control arm	50% at 3 years	● 40% ● 60%	
Number of progression events for final analysis	281 per arm		Stage 2 analysis conducted when 281 events occurred in the control arm

Results

***Table 2**. Expected trial duration under different scenarios.*

Scenario *	Expected time of stage 2 analysis (years)	Expected time of stage 1 analysis (years)	% of total participants by stage 1 analysis**
Base-case	6.1	3.4	58%
Recruitment 20% slower	6.5	3.7	51%
Recruitment 20% faster	5.7	3.3	66%
2 experimental arms in stage 2	6.2	3.4	58%
3 experimental arms in stage 2	6.3	3.4	58%
45% progression rate at 3 years	6.6	3.4	58%
55% progression rate at 3 years	5.7	3.4	58%

* Assuming base-case parameters, except for the one mentioned

** = number of participants recruited in trial at the time of the interim analysis / total trial size (assuming all arms continued into stage 2)

Under the base-case scenario it is estimated that each arm would have recruited 350 participants by the time the Stage 1 analysis is conducted. If two arms then continue into stage 2 (one experimental and the control), each recruiting 250 further participants, a total of 1,900 participants would have been recruited into the trial.

Timeline figure:

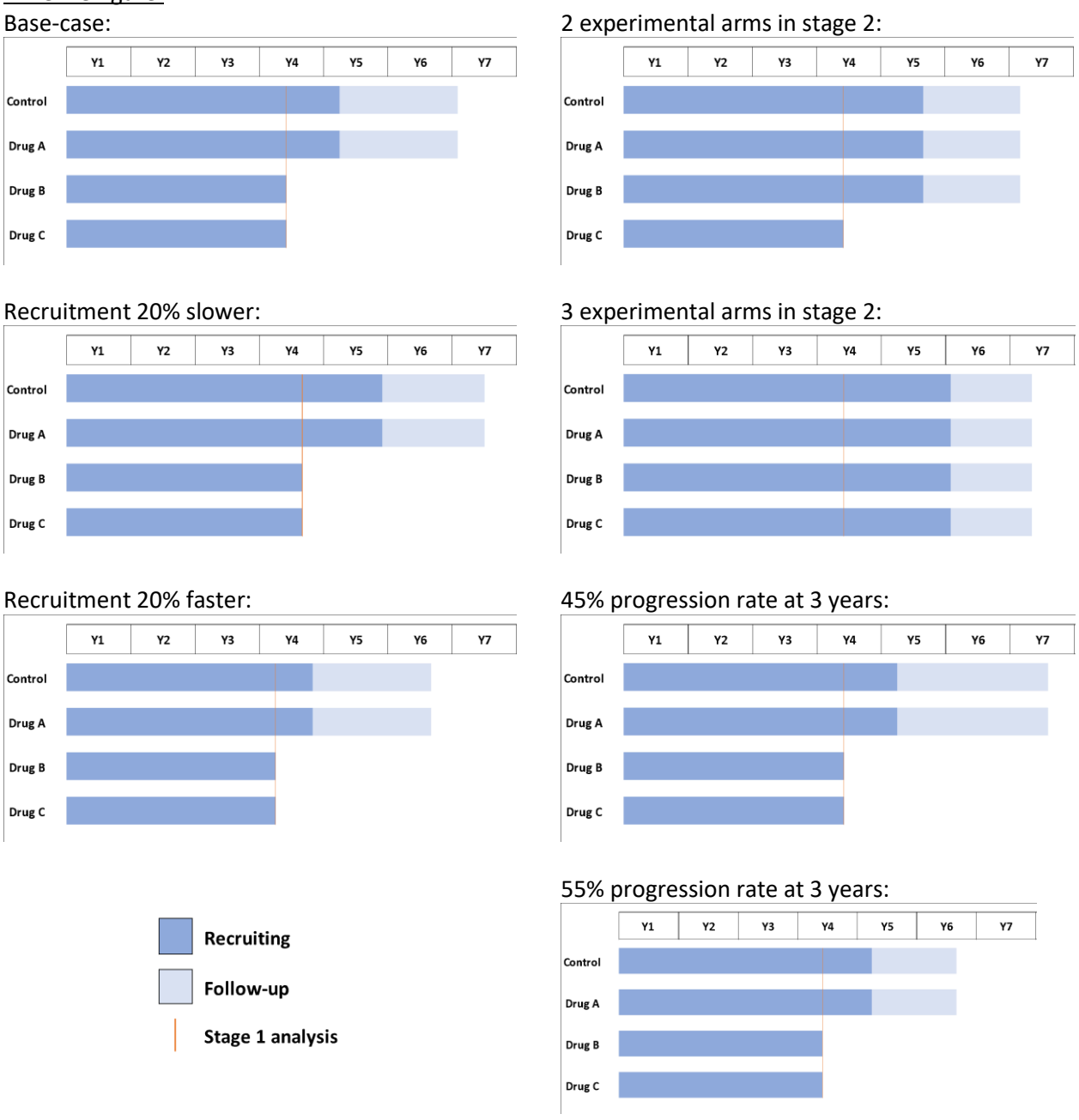


Figure 1. Schematic representation of trial timeline under different scenarios

eAppendix 5 – Trial simulations

The simulations aimed to assess the operating characteristics (e.g. overall type I and type II error rate) of the proposed trial design under different scenarios.

Methods

Data generation:

We recreated patient-level data aiming to replicate the proposed trial design. The different parameters used in the simulations are summarised in Table 1 below.

Outcomes
We generated two random correlated outcomes, one representing the stage-1 (interim analysis) outcome (brain atrophy rate), and the other the stage 2 (final analysis) outcome (time to clinical progression).

We drew correlated non-normally distributed outcomes by applying the following approach:

1. drawing two random correlated values from a bivariate normal distribution;
2. converting the normal draw to correlated uniforms, by applying the cumulative normal distribution function; and
3. converting the uniform draws to the desired normal and survival-time variables, by applying the inverse of the respective cumulative distribution functions.

The interim outcome was assumed to be normally distributed, while the final outcome was a time-to-event outcome, following an exponential distribution. The distributions are shown in more details in Table 1.

Sample size and censoring
We first simulated the two outcomes (brain atrophy and time to progression) for the total sample size (n=540 per arm). This represents the 'effective' sample size for the final analysis, after considering drop-out.
For the interim analysis, we used a randomly selected subset of the data equal to the planned number of observations (n_{int}=111 per arm, corresponding to the effective sample size for the interim analysis).
For the final analysis, we generated a censored time to event outcome from full data on time to progression by censoring all events that occurred after the time at which 281 events had accrued in the control arm. This represents the end of the trial, where the follow-up ends, and the final analysis is conducted. Those who have not yet progressed are censored.

Scenarios
We first look at the characteristics for a single experimental arm under H_0 (the treatment is ineffective) and under H_1 (the treatment is effective). For the overall trial characteristics, we considered four different scenarios: that 0, 1, 2 or 3 of the three experimental arms included an effective treatment (generated under H_1), and that the others were ineffective (generated under H_0). A set of 10,000 simulations was performed under each of the scenario.

All simulations were performed in Stata version 15.

Assessment:

For each simulated trial, we saved the results of the interim and final analysis (mean difference and hazard ratio, with 95% confidence intervals and p-values). We then summarised the findings by looking for each experimental arm whether they would have:

i) stopped at interim analysis, or
ii) continued to stage 2 but found ineffective at the final analysis, or
iii) continued to stage 2 and found to be effective at the final analysis.

Table 1. *Summary of simulation parameters.*

Parameter	Value	Note
Design		
Number of experimental arms	3	
Allocation ratio	1:1:1:1	=same number of patients in each arm
Number of effective treatments	0, 1, 2 or 3	Simulations were conducted for 0 to 3 effective treatment arms. An effective treatment had an effect size at interim and final analysis as stated below under H_1.
Interim analysis (stage 1)		
Outcome distribution	Normal(μ,σ^2)	Representing whole brain volume atrophy rate from baseline to 18 months, in % reduction per year
Sample size	111	Effective sample size per arm (before drop-out inflation), corresponding to 95% power per pairwise comparison
Effect size under H_0	0	
Effect size under H_1	0.15	Mean difference in atrophy rate under H_1
σ (Standard deviation)	0.55	
Alpha (1-sided)	0.35	Treatment continues to second stage if one sided p-value≤0.35
Analysis	Linear regression	Pairwise comparison of each active arm to the control arm, to estimate the mean difference.
Final analysis (stage 2)		
Survival distribution	Exponential(λ)	Representing the time from baseline to clinical progression
λ	0.231	Rate parameters in control arm. Corresponding to a probability of progression of 50% at 3 years.
Sample size	540	Effective sample size per arm, before drop-out inflation
Number of events	281	Follow-up censored after 281 events occurred in the control arm, for 90% power per pairwise comparison
Hazard ratio under H_0	1.0	
Hazard ratio under H_1	0.75	Expected treatment effect on time to progression under H1
Alpha (1-sided)	0.025	Experimental treatment declared superior if one-sided p-value <0.025. Note that this is consistent with the standard two-sided alpha of 0.05, but we focus here on whether the experimental arm is found superior.
Analysis	Cox regression	Pairwise comparison of each active arm to the control arm, to estimate the hazard ratio.
Correlation	0.5	This is the correlation between interim and final outcome (used in step 1 described in methods). Different values had little effect on overall trial characteristics.

Notes: Sample size are per arm. H0=null hypothesis, assuming treatment has no effect on brain atrophy or time to progression. H1=alternative hypothesis, assuming treatment is effective on brain atrophy and time to progression.

Results

Individual arm characteristics

Based on 10,000 simulations, we found that under the null hypothesis (treatment genuinely has no effect on the interim and final outcome), a treatment has 1.5% chance to be declared effective (being continued into the 2nd stage and obtaining a one sided p-value<0.025 at the final analysis). Under the alternative hypothesis, an effective treatment has 86.8% chance to be declared effective at the final analysis. These are lower than the standard 2.5% type I error rate and 90% power, because of the additional chance to stop treatment at the interim analysis without rejecting the null hypothesis.

Table 2. *Probability of different outcomes for pairwise comparisons under null and alternative hypothesis*

	H_0 (Ineffective treatment)	**H1 (Effective treatment)**
Stopped at stage 1 analysis	65.0%	5.0%
Continued into stage 2, but not found significantly superior at final analysis	33.5%	8.2%
Continued into stage 2, and found superior at final analysis	1.5%	86.8%

Results are column %, based on 10,000 simulations.

Overall trial characteristics

Table 3 shows the overall trial characteristics, depending on the number of truly effective treatments that enter the trial. In all scenarios, the probability of wrongly concluding that one or more treatment(s) are effective (family-wise type I error) is below 3.7%. The chance of correctly concluding that at least one treatment is effective (power) if a single effective drug enters the trial is around 87.3%, but this increases to above 96% if more than one effective drug enters the trial.

Table 3. *Trial operating characteristics according to number of effective treatments entering the trial.*

		Number of (truly) effective treatments at start of trial			
		0	**1**	**2**	**3**
Number of experimental arms in 2nd stage	0	40.5%	4.3%	1.0%	0.4%
	1	26.8%	45.8%	6.4%	2.3%
	2	19.3%	30.7%	58.8%	9.2%
	3	13.4%	19.2%	33.8	88.1%
Detected at least 1 truly effective ('power')		N/A	87.3%	96.2%	98.3%

At least 1 ineffective found significant ('type I error')	3.7%	2.6%	1.3%	N/A

Results are column %, based on 10,000 simulations for each of the four scenarios (number of truly effective treatments). N/A= Not applicable.

Monte Carlo error (uncertainty in the simulation estimates) is between 0.13% and 0.33% for the power, and between 0.11% and 0.19% for the type I error.