Berry Consultants
Statistical Innovation

Adaptive Trial Design for the clinical trial: Optimal Delay Time to Initiate Anticoagulation after Ischemic Stroke in Atrial Fibrillation (START)

Submitted to Seton
By Berry Consultants, LLC
January 28, 2018

1.0 OVERVIEW

This study is a multi-center, prospective, randomized, pragmatic trial to evaluate the optimal delay time to initiate anticoagulation with a non-vitamin K oral anticoagulant (NOAC) after ischemic stroke in patients with non-valvular atrial fibrillation. The study will include two populations, a mild/moderate stroke group and a severe stroke group. Separate trials will be run within each population. In the mild/moderate stroke population, we will randomize 1,000 subjects among 4 arms of time to anticoagulation treatment start time windows. The four arms are: Day 3 (48-72 hours), Day 6 (120-144 hours), Day 10 (216-240 hours) and Day 14 (312-336 hours). In the severe stroke population, we will randomize 500 subjects among the following 4 time to anticoagulation treatment start time arms: Day 6 (120-144 hours), Day 10 (216-240 hours), Day 14 (312-336 hours), and Day 21 (480-504 hours). The primary endpoint for both trials is a dichotomous composite outcome that includes any of the following within 30 days of the index stroke: Ischemic Events (symptomatic ischemic stroke or system embolism) or Hemorrhagic Events (symptomatic hemorrhagic transformation of index ischemic stroke, or other symptomatic intracranial hemorrhage, or major system hemorrhage).

The goal for both trials is to identify if one of the trial arms is superior to the other arms, i.e. select a time-to-treatment window, or to identify if an arm is inferior to the other arms, i.e. eliminate a window from standard practice.

The ischemic and hemorrhagic events within the composite primary endpoint will be modeled separately using their expected monotonic property that the risk of an event increases (ischemic) or decreases (hemorrhage) as the time-to-treatment interval lengthens. This trial will include frequent interim analyses where the primary outcome will be analyzed and new randomization probabilities will be calculated to favor the arms that have a better risk profile.

This document describes the features of the adaptive design, including the statistical models, decision rules, and simulation scenarios explored in the simulations. All simulations were conducted using FACTS (Fixed and Adaptive Clinical Trial Simulator, Berry Consultants, Austin, TX) software. FACTS version 6.1.1 was used for these simulations.

2.0 PRIMARY COMPOSITE ENDPOINT AND PRIMARY ENDPOINT ANALYSIS

The primary endpoint for both trials is a composite of two component events, ischemic and hemorrhagic events. Events for each trial will be modeled separately and the same models will be used for both trials.

2.1. *PRIMARY ENDPOINT COMPONENT – ISCHEMIC EVENTS*

The primary endpoint component for ischemic events is dichotomous and is measured at 30 days after the index stroke. A response corresponds to a negative outcome for the subject.

2.1.1. Ischemic Events Dose Response Model

Let Y_i be the ischemic events dichotomous outcome measured at day 30 for the *ith* subject. Label the four treatment arms (referred to as doses) as d = 1, 2, 3, 4. We model the outcomes as

$$Y_i \sim \text{Bernoulli}(p_{d_i})$$

where p_d is the underlying event rate for arm d. We transform the event rates onto the log-odds scale:

$$\theta_d = \log\left(\frac{p_d}{1-p_d}\right).$$

We model the dose-response for the primary endpoint using a sigmoidal (E_{max}) model:

$$\theta_d = a_1 + \frac{(a_2 - a_1)v_d^{a_4}}{v_d^{a_4} + a_3^{a_4}} \qquad d \in \{1, \ldots, 4\},$$

where v_d is the effective dose strength. We consider the arms, i.e. doses, to be equally spaced ($v_1 = 1, v_2 = 2, v_3 = 3, v_4 = 4$).

These parameters are modeled with prior distributions:

$$a_1 \sim N(-3.5, 1^2)$$

$$a_2 \sim N(0.1, 0.1^2),$$

$$a_3 \sim N^+(2.5, 5^2),$$

and

$$a_4 \sim N^+(1, 5^2).$$

The notation $N^+(\mu, \sigma^2)$ refers to a positively truncated normal distribution. The slope of the sigmoid curve is the parameter a_2. This parameter has a prior mean of 0.1 and a prior standard deviation of 0.1, reflecting the prior belief that the rate of ischemic events is increasing as the time-to-start is increased.

2.2. *PRIMARY ENDPOINT COMPONENT – HEMORRHAGIC EVENTS*

The primary endpoint component for hemorrhagic events is dichotomous and is measured at 30 days after the index stroke. A response corresponds to a negative outcome for the subject.

2.2.1. Hemorrhagic Events Dose Response Model

Let X_i be the second component of the dichotomous outcome, i.e. hemorrhagic events, measured at 30 days for the *ith* subject. We model the outcomes as:

$$X_i \sim \text{Bernoulli}(q_{d_i})$$

where q_d is the underlying event rate for arm d. We transform the event rates onto the log-odds scale:

$$\xi_d = \log\left(\frac{q_d}{1-q_d}\right).$$

We model the dose-response for the hemorrhagic component using a monotonic normal dynamic linear model (NDLM) as follows. The first dose in the model has prior:

$$\xi_1 \sim N(-2.94, 1^2).$$

Subsequent doses have a prior based on the estimate of the response on the preceding dose:

$$\xi_d \sim \text{N}^-(\theta_{d-1}, \tau_{d-1}^2) \qquad d \in \{2,3,4\},$$

where $\text{N}^-(\mu, \sigma^2)$ is the negative truncated normal distribution.

The variance in the prior dose is based on τ^2, the underlying dose to dose variance. The drift parameter (variance component) τ^2 dictates the amount of smoothing from dose to dose in the model. Small values of τ indicate that the responses of successive doses are likely to be very close, and that there is more borrowing (smoothing). It has prior distribution:

$$\tau^2 \sim IG(0.25, 0.0625).$$

2.3. *Primary Composite Endpoint Utility Function*

The adaptive aspects of the trial will be based on a utility function. The utility is a unit-less function of the 2 endpoint components, constructed by specifying the 1-dimensional component for each component and then combining them. The primary component of the utility function is the ischemic event and the secondary component is the hemorrhagic event. The primary and secondary components of the utility are defined as:

$$U_1(p_d) = w_1 * p_d \; where \; w_1 = -1$$

and

$$U_2(q_d) = w_2 * q_d \; where \; w_2 = -1.$$

These utility is the negative of the sum of the event rates:

$$U(p_d, q_d) = \; U_1(p_d) + U_2(q_d) = -[p_d + q_d].$$

Hence the trial design is designed to minimize the sum of the ischemic and hemorrhagic event rates.

2.4. *Interim Analysis to Update Utility Function Weights*

At one time point during each trial, a committee will meet to evaluate the long-term outcomes for subjects that had ischemic or hemorrhagic events. The committee will consider whether the weights of the outcomes should be changed because the long-term outcome indicates subjects with either ischemic or hemorrhagic outcomes had worse long-term outcomes. Based on the results of this review, the weighting of the ischemic and hemorrhagic events may be changed. Currently the weights are -1 for both outcomes, i.e. equal weighting, and these are the weights used in the simulations that follow. For the mild/moderate population, this review will occur after 400 subjects have been randomized. For the severe population, this review will occur after 200 subjects have been randomized.

2.5. *Final Analysis*

At the end of the trial we will calculate the probability that each arm is the best arm based on the utility function described above. The trial goal is to identify if one of the trial arms is better than the other arms, i.e. select a time-to-treatment window, or to identify if an arm is inferior to the other arms, i.e. eliminate a window from the standard practice. An arm is selected as being statistically the best arm if:

$$\Pr(\text{Arm } d \text{ is best}) > 0.75.$$

An arm is designated as statistically inferior to the other arms if:

$$\Pr(\text{Arm } d \text{ is best}) < 0.01.$$

3.0 ADAPTIVE DESIGN

3.1. *Interim Analysis Schedule*

Interim analyses will occur after every 100 subjects are randomized, starting with the 100th subject randomized in both trials. For the mild/moderate stroke trial, interim analyses will end after the 900th subject is randomized. For the severe stroke trial, interim analyses will end after the 400th subject is randomized.

3.2. *Response Adaptive Randomization*

Up to the first interim analysis, subjects will be equally randomized in blocks of size 8 with ratio 2:2:2:2. After this initial "burn-in" period, response adaptive randomization will begin, with the goal of preferentially allocating subjects to the time-to-treatment arms that are more promising. Once adaptive allocation begins, subjects will be randomized with allocation probabilities weighted according to the variance components:

$$V_d = \sqrt{\frac{\Pr(d = d_{Umax})\, Var(U_d)}{n_d + 1}}$$

where Pr($d = d_{Umax}$) is the probability that dose d is the dose with highest utility, Var(U_d) is the posterior variance of the utility at dose d, and n_d is the current number of subjects allocated to dose d. The randomization probabilities will be updated at each interim. They will be weighted according to the V_d and the weights will be normalized to sum to 1.

The design does not include any early stopping or dropping of time-to-treatment arms. However, the design does include a *temporary* arm dropping feature. That is, if a randomization allocation probability drops to less than 0.10 then that arm's allocation probability is set to zero for the next 100 subjects randomized. The study arm will remain in the trial, and the allocation probability will be updated again at the next interim analysis, and if the probability is above 0.10 the arm will have positive randomization probabilities.

4.0 OPERATING CHARACTERISTICS

We simulate the design as described above, separately for the mild/moderate stroke trial and the severe stroke trial to determine its operating characteristics. We consider various null hypothesis scenarios and alternative hypothesis scenarios. To simulate the design, we simulate virtual patients, conduct the trial as specified based on the outcomes observed for the virtual patients, and record the results. We repeat this process 10,000 times for each scenario and report the average results.

4.1. *Simulation Scenarios Mild/Moderate Stroke Trial*

We consider 10 scenarios in order to characterize the trial's operating characteristics (Table 1). The scenarios include rates of events between 5% and 10% for the combined ischemic and hemorrhage events deemed plausible for this subject population. In addition, we included scenarios where the rates are lower than estimated, that is, the rates are half that originally expected, i.e. combined rates between 2.5% and 5%.

Scenarios 1 and 6 are situations where one arm is the most beneficial, specifically the longest time-to-treatment window of Day 14, and the other arms are all similar in the rate of composite events observed. Scenarios 2 and 7 are situations where

one arm is clearly the least beneficial, specifically the shortest time-to-treatment window of Day 3. Two null scenarios, scenarios 3 and 8, are presented to evaluate the trial performance when there is no difference in the composite outcome rates between the four arms. Scenarios 4 and 9 vary the assumption of the monotonic property across the arms, and designate a middle arm, time-to-treatment window of Day 10 as most beneficial. Finally, scenarios 5 and 10 consider a situation where there are two similar beneficial arms, the shortest time-to-treatment window and the longest time-to-treatment window.

For all these simulation scenarios for the mild/moderate population, we assume an accrual rate of 200 subject per year, i.e. 3-4 subjects per week. The summaries below considered an accrual rate of 3 subjects per week.

Table 1.

Scenario	**Event Rates**		**Arm 1 Day 3**	**Arm 2 Day 6**	**Arm 3 Day 10**	**Arm 4 Day 14**
1	Day14best	ischemic	0.020	0.020	0.030	0.030
		hemorrhagic	0.080	0.080	0.080	0.020
		composite	0.100	0.100	0.110	0.050
2	Day3worst	ischemic	0.020	0.020	0.030	0.030
		hemorrhagic	0.080	0.060	0.040	0.020
		composite	0.100	0.080	0.070	0.050
3	None	ischemic	0.020	0.040	0.060	0.080
		hemorrhagic	0.080	0.060	0.040	0.020
		composite	0.100	0.100	0.100	0.100
4	InvertU-Day10	ischemic	0.010	0.020	0.030	0.100
		hemorrhagic	0.120	0.090	0.020	0.010
		composite	0.130	0.110	0.050	0.110
5	Day3and14better	ischemic	0.020	0.040	0.060	0.080
		hemorrhagic	0.080	0.080	0.080	0.020
		composite	0.100	0.120	0.140	0.100
6	Day14best (half)	ischemic	0.010	0.010	0.015	0.015
		hemorrhagic	0.040	0.040	0.040	0.010
		composite	0.050	0.050	0.055	0.025
7	Day3worst (half)	ischemic	0.010	0.010	0.015	0.015
		hemorrhagic	0.040	0.030	0.020	0.010
		composite	0.050	0.040	0.035	0.025
8	None (half)	ischemic	0.010	0.020	0.030	0.040
		hemorrhagic	0.040	0.030	0.020	0.010
		composite	0.050	0.050	0.050	0.050
9	InvertU-Day10 (half)	ischemic	0.005	0.010	0.015	0.050
		hemorrhagic	0.060	0.045	0.010	0.005
		composite	0.065	0.055	0.025	0.055
10	Day3and14better (half)	ischemic	0.010	0.020	0.030	0.040
		hemorrhagic	0.040	0.040	0.040	0.010
		composite	0.050	0.060	0.070	0.050

4.2. *Simulation Results Mild/Moderate Stroke Trial*

Table 2 presents operating characteristics for the ten scenarios. For each scenario, the table summarizes the mean sample size per arm, the probability that each arm is selected as the *best* and the probability that each arm is selected as the *worst.*

Table 2.

Scenario		Arm 1 Day 3	Arm 2 Day 6	Arm 3 Day 10	Arm 4 Day 14	
1	Day14best	104.2	143.7	217.5	534.7	*N*
		0	0	0	0.949	*Pr. Max>0.75*
		0.895	0.744	0.466	0	*Pr. Max<0.01*
2	Day3worst	120.6	175.6	270.1	433.7	*N*
		0	0	0.002	0.614	*Pr. Max>0.75*
		0.788	0.392	0.023	0.000	*Pr. Max<0.01*
3	None	233.0	244.5	255.0	267.4	*N*
		0.006	0.009	0.018	0.018	*Pr. Max>0.75*
		0.076	0.008	0.004	0.071	*Pr. Max<0.01*
4	InvertU-Day10	102.1	202.2	474.6	221.1	*N*
		0	0	0.971	0	*Pr. Max>0.75*
		0.955	0.613	0	0.652	*Pr. Max<0.01*
5	Day3and14better	231.8	220.9	217.3	330.0	*N*
		0.016	0.001	0.000	0.226	*Pr. Max>0.75*
		0.075	0.040	0.037	0.016	*Pr. Max<0.01*
6	Day14best (half)	112.7	167.7	246.3	473.3	*N*
		0	0	0	0.852	*Pr. Max>0.75*
		0.84	0.50	0.13	0	*Pr. Max<0.01*
7	Day3worst (half)	123.4	185.8	271.9	418.9	*N*
		0	0	0.002	0.543	*Pr. Max>0.75*
		0.760	0.263	0.006	0.000	*Pr. Max<0.01*
8	None (half)	201.2	232.9	260.9	305.1	*N*
		0.000	0.002	0.015	0.034	*Pr. Max>0.75*
		0.131	0.012	0.001	0.037	*Pr. Max<0.01*
9	InvertU-Day10 (half)	127.0	219.6	371.4	282.0	*N*
		0	0	0.747	0.002	*Pr. Max>0.75*
		0.802	0.260	0	0.260	*Pr. Max<0.01*
10	Day3and14better (half)	198.2	219.7	241.7	340.4	*N*
		0.001	0	0.002	0.184	*Pr. Max>0.75*
		0.130	0.034	0.006	0.010	*Pr. Max<0.01*

In the scenario where the longest time-to-treatment window of Day 14 was the best arm with a composite rate of 5%, scenario 1, the mean number of subjects randomized to the Day 14 arm was 535. Only an average of 218, 144, and 104 subjects were randomized to the decreasing time-to-treatment windows with rates of 11%, 10% and 10% respectively. The Day 14 arm was chosen as the best arm with a high probability, 0.95, and the Day 3 arms was designated statistically inferior with a high probability of 0.90 in the simulated trials. If the event rates are halved (5%, 5%, 5.5%, 2.5%), scenario 6, a higher proportion of subjects is still randomized to the Day 14 arm. The Day 14 arm is selected as best with a lower probability of 0.85 and the Day 3 arm is designated as inferior with a probability of 0.84.

For the scenarios where one arm was inferior compared to the others, scenario 2, only 121 subjects were randomized to the Day 3 arm which had a simulated rate of 10%. For the Day 6, 10, and 14 arms, with rates of 8%, 7% and 5%, the average number of subjects allocated to the arms was 176, 270, and 434 respectively. The Day 3 arm is identified as the statistically inferior arm with a probability of 0.79. If the assumed event rates are halved, scenario 7, the Day 3 arm is identified as the statistically inferior arm with a probability of 0.76.

Considering scenarios where there is no difference in the rates among the arms, scenarios 3 and 8, the allocation of subjects is similar across the four arms. An arm is designated as the best less than 2% for both scenarios. The designation of an arm as being the worst is at most 8% for scenario 3, the higher rate scenario, and at most 13% for the lower rate scenario, both low probabilities.

In situations where the better performing arm is a middle arm, Day 10 (scenario 4) is the better performing arm, more subjects are allocated to that arm, 475, compared to the other arms. In addition, that arm is selected as best with a high probability of 0.97. Also, in scenario 4, the trial is able to identify the Day 3 arm as statistically inferior with a probability of 0.96. For scenario 9, the probabilities are not as high. This is a result of less separation between the arms. In scenario 4 the composite rates across arms are 13%, 11%, 5%, and 11% respectively. However, in scenario 9, the simulated composite rates are 6.5%, 5.5%, 2.5%, and 5.5% respectively.

In the scenarios where two arms are similarly the better arms, scenarios 5 and 10, the trial does not have a high probability of declaring one arm as better. This is expected because the trial is designed to be able to differentiate if one arm is the better performing arm, or to identify one arm as the worst performing arm.

In general, for all scenarios, the response adaptive randomization allows for improved treatment of the subjects in the trial by placing them on more effective arms.

If we have a situation where event rates are lower than estimated, for example, halved, the probabilities of selecting an arm as better, or designating as statistically

inferior were lower. The smaller number of events, and relative smaller difference in rates between arms in the scenarios simulated results in less differentiation in the probabilities an arm is the best.

4.3. *Simulation Scenarios Severe Stroke Trial*

We consider 5 scenarios in order to characterize the trial's operating characteristics (Table 3). The scenarios include rates of events between 12% and 26% for the combined ischemic and hemorrhage events deemed plausible for this subject population.

Scenario 1 is a situation where one arm is the most beneficial, specifically the longest time-to-treatment window of Day 21, and the other arms are all similar in the rate of composite events observed. Scenario 2 is a situation where one arm is clearly the least beneficial, specifically the shortest time-to-treatment window of Day 6. The null scenario, scenario 3, is presented to evaluate the trial performance when there is no difference in the composite outcome rates between the four arms. Scenario 4 varies the assumption of the monotonic property across the arms, and designate a middle arm, time-to-treatment window of Day 10 as most beneficial. Finally, scenario 5 considers a situation where there are two similar beneficial arms, the shortest time-to-treatment window and the longest time-to-treatment window.

For all the simulation scenarios, we assume an accrual rate of 1.15 subjects per week, approximately 60 subjects per year, with a maximum of 500 subjects enrolled.

Table 3.

Scenario	**Event Rates**		**Arm 1 Day 6**	**Arm 2 Day 10**	**Arm 3 Day 14**	**Arm 4 Day 21**
1	Day14best	ischemic	0.040	0.040	0.060	0.060
		hemorrhagic	0.160	0.160	0.140	0.060
		composite	0.200	0.200	0.200	0.120
2	Day3worst	ischemic	0.040	0.040	0.060	0.060
		hemorrhagic	0.180	0.140	0.100	0.060
		composite	0.220	0.180	0.160	0.120
3	None	ischemic	0.040	0.080	0.120	0.160
		hemorrhagic	0.180	0.140	0.100	0.060
		composite	0.220	0.220	0.220	0.220
4	InvertU-Day10	ischemic	0.040	0.060	0.080	0.180
		hemorrhagic	0.160	0.080	0.080	0.040
		composite	0.200	0.140	0.160	0.220
5	Day3and14better	ischemic	0.040	0.080	0.120	0.160
		hemorrhagic	0.160	0.160	0.140	0.060
		composite	0.200	0.240	0.260	0.220

4.4. *SIMULATION RESULTS MILD/MODERATE STROKE TRIAL*

Table 4 presents operating characteristics for the ten scenarios. For each scenario, the table summarizes the mean sample size per arm, the probability that each arm is selected as the *best* and the probability that each arm is selected as the *worst.*

Table 4.

Scenario		Arm 1 Day 6	Arm 2 Day 10	Arm 3 Day 14	Arm 4 Day 21	
1	Day14best	73.7	86.2	118.8	221.4	*N*
		0.000	0.000	0.000	0.771	*Pr. Max>0.75*
		0.603	0.398	0.103	0.000	*Pr. Max<0.01*
2	Day3worst	71.7	90.3	130.1	208.0	*N*
		0.000	0.000	0.001	0.573	*Pr. Max>0.75*
		0.639	0.300	0.017	0.000	*Pr. Max<0.01*
3	None	126.5	119.6	117.8	136.1	*N*
		0.016	0.007	0.005	0.024	*Pr. Max>0.75*
		0.039	0.009	0.008	0.059	*Pr. Max<0.01*
4	InvertU-Day10	119.4	144.8	123.4	112.5	*N*
		0.001	0.138	0.008	0.001	*Pr. Max>0.75*
		0.088	0.000	0.008	0.274	*Pr. Max<0.01*
5	Day3and14better	134.8	114.0	107.0	144.2	*N*
		0.059	0.001	0.000	0.065	*Pr. Max>0.75*
		0.013	0.010	0.024	0.052	*Pr. Max<0.01*

In the scenario with the longest time-to-treatment window of Day 21 was the best arm with a composite rate of 12%, scenario 1, the mean number of subjects randomized to the Day 14 arm was 222. Only an average of 119, 87, and 74 subjects were randomized to the decreasing time-to-treatment windows with rates of 20% for each of those arms. The Day 21 arm was chosen as the best arm with a probability, 0.77, and the Day 3 arm was designated statistically inferior with a probability of 0.60 in the simulated trials.

For the scenario where one arm was inferior compared to the others, scenario 2, only 72 subjects were randomized to the Day 6 arm which had a simulated rate of 10%. For the Day 10, 14, and 21 arms, with rates of 18%, 16% and 12%, the average number of subjects allocated to the arms was 91, 131, and 208 respectively. The Day 3 arm is identified as the statistically inferior arm with a probability of 0.64.

Considering a scenario where there is no difference in the rates among the arms, scenario 3, the allocation of subjects is similar across the four arms. An arm is designated as the best less than 2.5%. The designation of an arm as being the worst is at most 6%.

In situations where the better performing arm is a middle arm, Day 10 (scenario 4) is the better performing arm, more subjects are allocated to that arm, 145, compared to the other arms. However, with the rates across the arms having less separation, composite rate of 20%, 14%, 16% and 22%, the algorithm is not able to distinguish a best of worst arm as low probabilities are seen for both of these results. A similar situation was seen for scenario 5, where two arms are similarly the better arms. Again, the trial does not have a high probability of declaring one arm as better. This is expected because the trial is designed to be able to differentiate if one arm is the better performing arm, or to identify one arm as the worst performing arm.

In general, for all scenarios, the response adaptive randomization allows for improved treatment of the subjects in the trial by placing them on more effective arms.