Appendix M Bayesian Adaptive Design for a Study of the Sensei® Robotic System

Berry Consultants
Statistical Innovation

Bayesian Adaptive Design for a Study of the Sensei Robotic System

Submitted to Hansen Medical
June 24, 2013

1.0 Introduction

This document outlines the adaptive design for the Sensei Robotic System in atrial fibrillation patients. The purpose of this document is to provide a description of the design along with details of the statistical models and simulation results.

The primary efficacy endpoint is freedom from failure at 12 months, after a 3-month blanking period. The primary safety endpoint is freedom from Major Complications within 7 days of the ablation and esophageal injury or pulmonary vein stenosis through day 180.

The trial is a single-arm study, comparing the efficacy and safety of the device against Objective Performance Criteria (OPC).

The goal of the trial is to determine if robotic placement of the catheter meets the OPC goals for BOTH the primary efficacy and primary safety endpoints. An adaptive design is proposed which allows for early discontinuation of enrollment for either expected success or futility, based upon interim evaluations of the accumulating data and prospective rules defined here.

The sample size of the trial will be determined by the adaptive design. The trial will enroll a maximum of 250 patients. Interim monitoring will be conducted starting after the first 125 patients are enrolled, and continuing every 25 patients thereafter. At each interim analysis, enrollment may be stopped for expected success or for futility.

Up to three additional interim analyses will be conducted after full enrollment (0, 3, and 6 months). At these interims, trial success may be declared early if there is greater than 99.9% predictive probability of final success AND a minimum of 100 patients have complete safety data and 80 patients have complete efficacy data.

2.0 Statistical Modeling

2.1. Final Analysis

Let Y be an indicator of "success" (failure-free) at 12 months, where $Y = 1$ means success, and $Y = 0$ indicates failure. We model the outcomes as

$$Y \sim \text{Binomial}(n, \theta),$$

where θ is the failure-free rate and is modeled with prior Beta(1, 1), which is equivalent to 2 observations of weight on a 50% rate. The failure-free rate will be compared to an OPC of 54%.

Similarly, let X be an indicator of freedom from serious adverse events at 6 months, where $X = 0$ means that the patient experienced an AE, and $X = 1$ indicates freedom from primary safety endpoint events. We model the outcomes as

$$X \sim \text{Binomial}(n, \pi),$$

where π is the AE-free rate and is modeled with prior Beta(0.1, 0.1). We choose a distribution with the same mean, but a tenth of the prior weight compared to the prior for the efficacy endpoint. The AE-free rate will be compared to an OPC of 84%.

The trial will be considered a success if BOTH endpoints achieve at least 97.5% posterior probability of exceeding the OPC. The 97.5% level was selected to control experiment-wise type I error rate below 5%.

2.2 Longitudinal Model

At the time of each interim analysis, some patients will not have completed the full evaluation period. A longitudinal model will be employed to enable final observations to be imputed for those subjects with incomplete information.

For example, recently enrolled patients who are currently failure-free but have only been observed for a portion of the observation period will have "censored" final outcomes. We will use a time-to-event model to multiply impute final outcomes for each subject with partial data. These imputed outcomes are then used to update the posterior distributions of the primary efficacy and safety endpoints.

The longitudinal model for each endpoint is a piecewise exponential model with three segments. For consistency with the published results from the ThermoCool trial, we convert time to weeks and use intervals of (0,2], (2,8], and (8,39] for the time-to-failure model. In this model, time zero is defined as the end of the blanking period.

The failure times are modeled as:

$$T_{fail} \sim \text{PE}(\lambda_1, \lambda_2, \lambda_3).$$

The independent prior distributions are:

$$\lambda_1 \sim \text{Gamma}(5, 29.9),$$

$$\lambda_2 \sim \text{Gamma}(5, 694.4),$$

$$\lambda_3 \sim \text{Gamma}(5, 1190.5),$$

with mean hazard rates of 0.167, 0.007, and 0.004 events per week, respectively. These distributions are centered at the hazard rates reported in the ThermoCool SSED, but down-weighted so that they contribute 5 observations worth of weight.

The time-to-AE model uses different segments of (0,1], (1,4), (4,26], consistent with the expectation that the rate of adverse events may differ in the first week after the procedure, the next 3 weeks, and the remainder of the 6-month evaluation period.

The AE times are modeled as:

$$T_{AE} \sim \text{PE}(\gamma_1, \gamma_2, \gamma_3),$$

with independent prior distributions:

$$\gamma_1 \sim \text{Gamma}(1, 25),$$

$$\gamma_2 \sim \text{Gamma}(1, 50),$$

$$\gamma_3 \sim \text{Gamma}(1, 1000).$$

2.3 Bayesian Predictive Probabilities

We define the following two predictive probabilities for use in decision rules for early stopping for expected success and futility:

- We let pp_{now} be the predictive probability that BOTH the efficacy analysis and safety analysis will be meet the success threshold if we stop enrollment at the current number of subjects and follow all enrolled subjects to the 12-month outcome.

- We let pp_{max} be the predictive probability that BOTH the efficacy analysis and safety analysis will meet the success threshold if we enroll to the maximum sample size and follow all subjects to the 12-month outcome.

3.0 Adaptive Design

3.1 Adaptive Sample Size

Interim monitoring will occur after 125, 150, 175, 200, and 225 subjects have been enrolled. At each interim analysis, the predictive probabilities for expected success and futility will be computed and compared to pre-specified early stopping criteria.

3.1.1 Expected Success

If there is at least 95% probability that BOTH the efficacy analysis and safety analyses will meet the success criteria if enrollment stops at the current sample size and all patients are followed to the 12-month endpoint, then enrollment will stop early for expected success. That is, stop enrollment for expected success if $pp_{now} > 0.95$.

3.1.2 Early Futility

If there is less than 1% probability that the BOTH the efficacy analysis and safety analyses will meet the success criteria when the maximum number of subjects is enrolled, the trial will stop early for futility. That is, stop for early futility if $pp_{max} < 0.01$.

3.2 Evaluation of Early Success

Up to three additional interims will be performed: after full enrollment and at 3 and 6 months after full enrollment to assess the criteria for an early declaration of success.

If there is at least 99.9% probability that BOTH the efficacy analysis and safety analyses will meet the success criteria after all enrolled subjects are followed to the 12-month endpoint, AND if a minimum of 80 and 100 patients have complete follow-up for the efficacy and safety endpoints, respectively, then the trial will declare early success.

3.3 Trial Completion

The final analysis will occur when both accrual and follow-up are complete for all subjects. If, at the completion of the trial, BOTH the efficacy and safety analyses are significant, then the trial will be a success. That is, the trial is successful if:

$$\Pr(\theta > 54\%) > 97.5\% \text{ and } \Pr(\pi > 84\%) > 97.5\%.$$

The 97.5% threshold was selected to control the experiment-wise type I error below 5%.

4.0 Example Trials

In this section, we present three example trials to illustrate the adaptive process. Each of the three example trials illustrates a different rate of accrual. The first example describes a trial that stops enrollment for expected success, and claims early success. The second example demonstrates early stopping for futility. The third example follows a trial that enrolls quickly.

4.1 Example Trial #1

The first interim analysis occurs when 125 subjects are enrolled, which occurs 81 weeks into the trial. The first row of Table 4.1 shows the available data. At this time, 58 subjects have complete data for the efficacy analysis, with 33 of those being failures. Thus, 67 subjects are ongoing with censored final outcome. Likewise, 91 subjects have complete data for the safety analysis, and 10 AEs have been observed.

We calculate the predictive probability that both endpoints will be successful if we stop enrollment now and follow all patients to the 12-month outcome. This predictive probability is 0.6308, which does not meet the criterion for early stopping. We also compute the predictive probability that both endpoints will be successful if enrollment continues to the maximum sample size. This probability is 0.8024. Since this probability is above 1%, enrollment continues to the next look at 150 subjects.

Table 4.1 Example Trial 1 interim results

N	Efficacy				Safety				Both	
	Complete	Fail	pp_{now}	pp_{max}	Complete	AEs	pp_{now}	pp_{max}	pp_{now}	pp_{max}
125	58	33	0.6456	0.8380	91	10	0.9792	0.9584	0.6308	0.8024
150	81	38	0.9940	0.9828	114	13	0.9656	0.9392	0.9600	0.9224
150	105	45	>0.9999	--	133	13	0.9996	--	0.9996	--

The next interim analysis occurs when 150 patients have been enrolled. At this time, the predictive probability of success with no additional enrollment has risen to 0.96, which exceeds the threshold to stop enrollment. However, this probability is not high enough to trigger a success claim. We continue to follow the enrolled patients, and conduct another analysis 3 months later.

At that time, the predictive probability of success is 0.9996, and the trial claims early success.

Figure 4.1 shows the available data and Kaplan-Meier plot for each endpoint at the time of the last analysis. Hash marks designate censored observations.

4.2 Example Trial #2

The trial begins by enrolling 125 subjects. At week 191, the first interim analysis is conducted. The predictive probability of trial success if enrollment continues to the maximum sample size is only 8.36%. Since this value is above the 1% threshold, enrollment precedes to the next interim at 150 subjects. At this time, final trial success is essentially impossible, and the trial stops for futility.

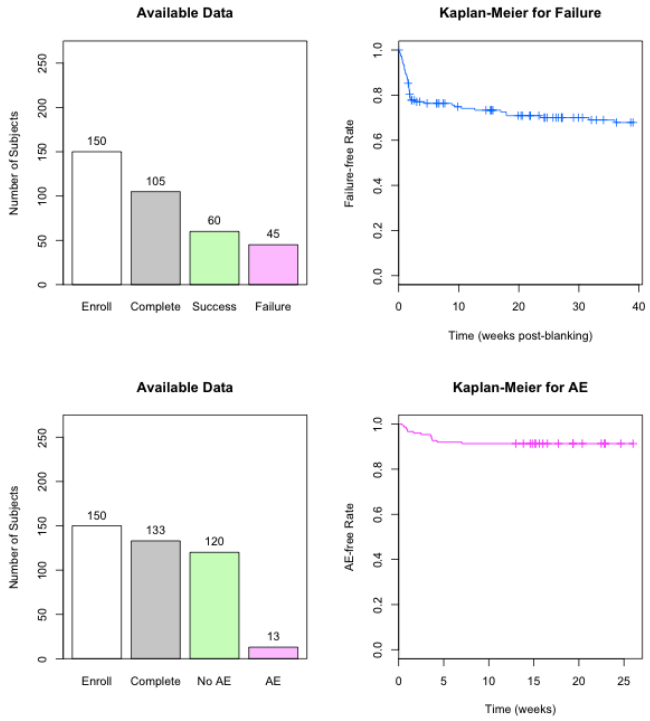

Figure 4.1: Example Trial 1. The available data and Kaplan-Meier plots at the time that the trial claims success. The top row corresponds to the efficacy data, and the bottom row is for the safety data.

The sequence of interim results is shown in Table 4.2. Kaplan-Meier curves are plotted in Figure 4.2.

Table 4.2 Example Trial 2 interim results

N	Efficacy				Safety				Both	
	Complete	Fail	pp_{now}	pp_{max}	Complete	AEs	pp_{now}	pp_{max}	pp_{now}	pp_{max}
125	88	15	>0.9999	>0.9999	105	18	0.0000	0.0836	0.0000	0.0836
150	112	18	>0.9999	>0.9999	137	28	0.0000	0.0000	0.0000	0.0000

We note that the efficacy analysis is highly likely to be successful. However, because the safety endpoint has low predicted probability of success, the predicted probability of success on both endpoints falls below the futility boundary.

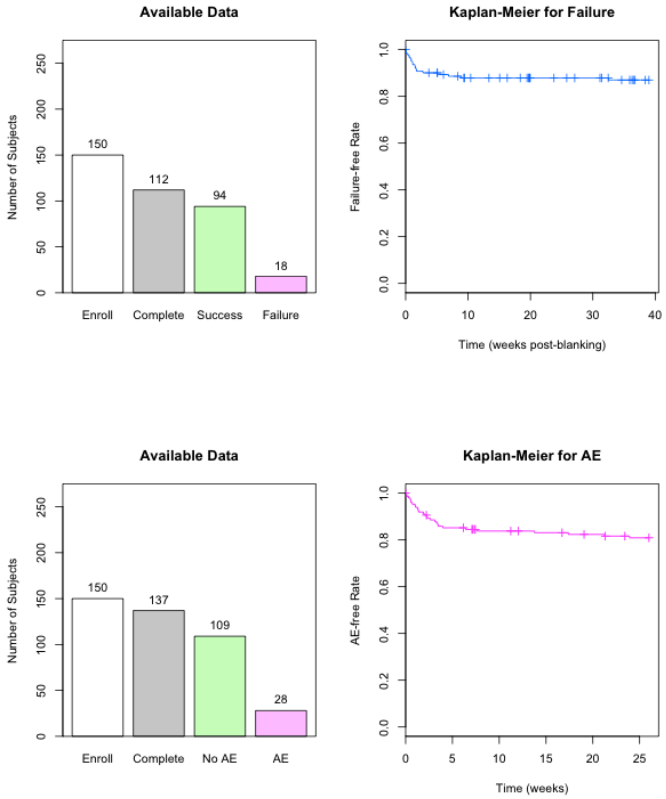

Figure 4.2: Example Trial 2. The available data and Kaplan-Meier plots at the time that the trial claims success. The top row corresponds to the efficacy data, and the bottom row is for the safety data.

4.3 Example Trial #3

In this trial, enrollment is very fast, and 125 subjects have enrolled by week 65. The data appears very positive for both endpoints, and the predictive probability of success exceeds the pre-specified threshold. The trial stops enrollment, but cannot immediately claim success.

Table 4.3 Example Trial 3 interim results

N	Efficacy				Safety				Both	
	Complete	Fail	pp_{now}	pp_{max}	Complete	AEs	pp_{now}	pp_{max}	pp_{now}	pp_{max}
125	32	21	0.9996	0.9992	74	8	0.9984	0.9908	0.9980	0.9900
125	63	27	>0.9999	>0.9999	96	8	>0.9999	>0.9999	>0.9999	>0.9999
125	88	32	>0.9999	>0.9999	125	8	>0.9999	>0.9999	>0.9999	>0.9999

Three months later, another interim analysis is performed. Again, the predictive probability of success is above 99.9%, but too few patients have complete follow-up data.

Another interim analysis occurs after another three months. At this analysis, the predictive probability of success continues to be above 99%, and there are 88 and 125

subjects with complete follow-up on the efficacy and safety endpoints, respectively, and the trial meets the criteria for early success.

The data and results for each interim are shown in Table 4.3 and Figure 4.3.

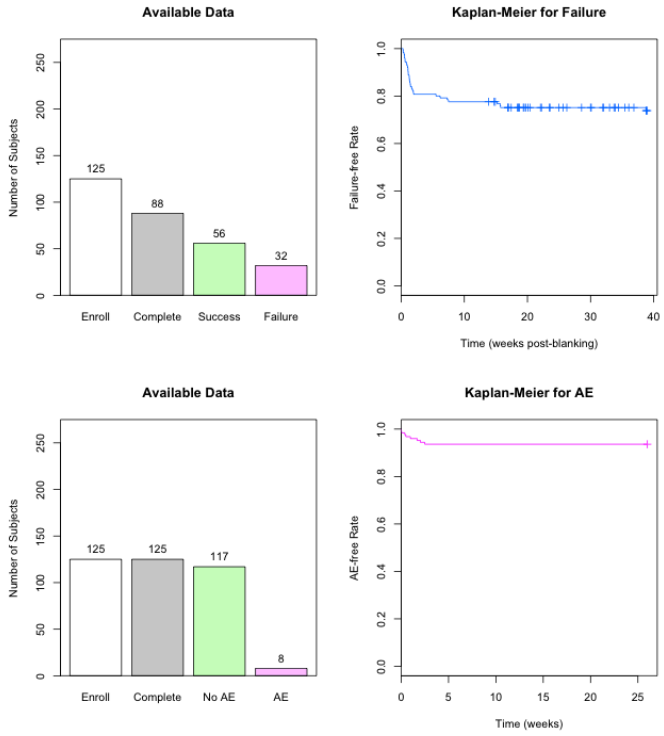

Figure 4.3: Example Trial 3. The available data and Kaplan-Meier plots at the time that the trial claims success. The top row corresponds to the efficacy data, and the bottom row is for the safety data.

5.0 Positive Simulation Scenarios and Operating Characteristics

In order to characterize and understand the performance of the design, we simulated the trial under a wide range of scenarios. In this section, we describe the simulation assumptions and results under various positive scenarios. In section 6.0, we investigate numerous null cases in order to assess type 1 error control.

5.1 Final Endpoint Profiles

We consider a wide variety of scenarios for each of the final endpoints. In order to have appropriate data at the interims, data is simulated from a piecewise exponential model, calibrated to the desired success rates. Thus, for each patient, we simulate whether the patient experiences a failure and the time at which the failure occurred. The profiles are constructed by taking advantage of the relationship:

$$\theta = \exp(-[2\,\lambda_1 + (8-2)\,\lambda_2 + (39-8)\,\lambda_3]).$$

From the ThermoCool SSED, we note that the reported hazard rates for the efficacy endpoint have the following pattern:

$$h_1 = (38.06)\,h_3$$

$$h_2 = (1.71)\,h_3$$

$$h_3 = (1)\,h_3,$$

for $h_3 = 0.004$.

The hazard rates for simulating data can then be computed by using these multiplication factors and specifying the desired success rate.

We use similar logic for calculating the hazard rates for the safety profiles. We assume multiplication factors of 50, 25, and 1 for γ_1, γ_2, and γ_3, respectively.

The resulting profiles are displayed in Tables 5.1 and 5.2.

Table 5.1 Efficacy Profiles

Failure-free Rate	Hazard Rate (λ)		
	(0, 2]	(2, 8]	(8, 39]
0.64	0.1447	0.0065	0.0038
0.66	0.1347	0.0061	0.0035
0.68	0.1250	0.0056	0.0033
0.70	0.1157	0.0052	0.0030
0.72	0.1065	0.0048	0.0028
0.74	0.0976	0.0044	0.0026

Table 5.2 Safety Profiles

AE-free Rate	Hazard Rate (γ)		
	(0, 1]	(1, 4]	(4, 26]
0.91	0.0321	0.0160	0.0006
0.92	0.0284	0.0142	0.0006
0.93	0.0247	0.0123	0.0005
0.94	0.0210	0.0105	0.0004

We cross all efficacy and safety profiles for a total of 24 positive scenarios. The set of null scenarios is described in section 6.0 of this report.

5.3. Accrual Rate

Patient entry to the trial is simulated from a Poisson process. The expected accrual rate is 6 subjects per month, following a 4-month ramp-up period. Additionally, we simulate accrual rates that are faster (12 per month) and slower (3 per month) than this expected rate.

5.4 Operating Characteristics

We simulate a total of 24 scenarios that combine the efficacy and safety profiles. Each of these scenarios was simulated under 3 accrual rates for 1000 trials each.

Table 5.3 summarizes the average sample size across the simulated trials, and the proportion of trials having outcomes of early success, late success, early futility, and late failure. The column "Early Success" indicates the proportion of trials for which a claim of success was made before complete follow-up of all patients.

To facilitate comparisons across tables, we highlight rows with the same safety rate in the same color.

Table 5.3 Positive Scenario Trial Outcomes (6 patients/month)

Efficacy Rate	Safety Rate	Mean N	Early Success	Late Success	Total Success	Early Futility	Late Failure
0.64	**0.91**	202.2	0.683	0.178	0.861	0.034	0.097
0.66	**0.91**	191.1	0.787	0.105	0.892	0.022	0.080
0.68	**0.91**	179.0	0.894	0.043	0.937	0.013	0.045
0.70	**0.91**	167.4	0.913	0.024	0.937	0.012	0.048
0.72	**0.91**	165.0	0.927	0.010	0.937	0.020	0.039
0.74	**0.91**	161.9	0.935	0.004	0.939	0.019	0.038
0.64	**0.92**	194.8	0.700	0.187	0.887	0.019	0.088
0.66	**0.92**	177.4	0.845	0.113	0.958	0.012	0.026
0.68	**0.92**	164.6	0.910	0.066	0.976	0.005	0.014
0.70	**0.92**	157.7	0.957	0.030	0.987	0.002	0.008
0.72	**0.92**	151.9	0.967	0.010	0.977	0.006	0.014
0.74	**0.92**	147.4	0.984	0.002	0.986	0.001	0.011
0.64	**0.93**	190.1	0.674	0.228	0.902	0.020	0.072
0.66	**0.93**	174.8	0.832	0.132	0.964	0.006	0.025
0.68	**0.93**	155.9	0.906	0.084	0.990	0.001	0.004
0.70	**0.93**	146.8	0.956	0.041	0.997	0	0.002
0.72	**0.93**	140.5	0.986	0.013	0.999	0	0
0.74	**0.93**	139.5	0.989	0.008	0.997	0	0.002
0.64	**0.94**	187.5	0.674	0.218	0.892	0.018	0.087
0.66	**0.94**	168.0	0.822	0.141	0.963	0.005	0.025
0.68	**0.94**	153.5	0.922	0.07	0.992	0	0.006
0.70	**0.94**	141.8	0.949	0.048	0.997	0.001	0
0.72	**0.94**	134.4	0.982	0.016	0.998	0	0
0.74	**0.94**	133.0	0.998	0.002	1	0	0

We evaluate power across a range of possible effect sizes for each endpoint. When the effect is large for both endpoints, the power is above 99%. For a moderate scenario in which the failure-free rate is 0.72 and the AE-free rate is 0.92, the power is 97.7%. The smallest effect considered here is an efficacy rate of 0.64. When paired with a safety rate of 0.91, the power is 86.1%.

For these scenarios, the average sample size ranges from approximately 133 to 202 subjects, with the smallest values occurring for the scenarios with large treatment effect on both endpoints. This behavior is expected due to the high probability of stopping enrollment for expected success when the effect size is large. The cumulative probability of stopping enrollment at each interim is shown in Tables 5.4 (expected success) and Table 5.5 (futility).

For the smallest effect of 0.64 for efficacy and 0.91 for safety, 58.9% of simulated trials stopped enrollment early for expected success at or before 225 subjects.

Table 5.4 Cumulative probability of stopping enrollment for expected success

Efficacy Rate	Safety Rate	Sample Size (N)				
		125	150	200	250	225
0.64	**0.91**	0.116	0.251	0.383	0.502	0.589
0.66	**0.91**	0.183	0.329	0.492	0.610	0.686
0.68	**0.91**	0.243	0.432	0.611	0.726	0.792
0.70	**0.91**	0.351	0.550	0.718	0.802	0.852
0.72	**0.91**	0.393	0.574	0.728	0.806	0.847
0.74	**0.91**	0.439	0.603	0.759	0.820	0.864
0.64	**0.92**	0.157	0.311	0.451	0.574	0.668
0.66	**0.92**	0.234	0.435	0.620	0.759	0.825
0.68	**0.92**	0.344	0.574	0.751	0.841	0.895
0.70	**0.92**	0.415	0.628	0.816	0.900	0.930
0.72	**0.92**	0.541	0.692	0.853	0.894	0.925
0.74	**0.92**	0.600	0.751	0.877	0.924	0.950
0.64	**0.93**	0.197	0.360	0.503	0.605	0.679
0.66	**0.93**	0.270	0.500	0.649	0.749	0.819
0.68	**0.93**	0.437	0.660	0.827	0.903	0.934
0.70	**0.93**	0.568	0.768	0.886	0.942	0.964
0.72	**0.93**	0.664	0.833	0.937	0.967	0.981
0.74	**0.93**	0.691	0.848	0.932	0.969	0.981
0.64	**0.94**	0.207	0.390	0.522	0.639	0.692
0.66	**0.94**	0.354	0.553	0.696	0.804	0.859
0.68	**0.94**	0.494	0.701	0.819	0.901	0.944
0.70	**0.94**	0.604	0.836	0.929	0.968	0.989
0.72	**0.94**	0.760	0.908	0.970	0.991	0.994
0.74	**0.94**	0.799	0.919	0.972	0.993	0.998

Berry Consultants
Statistical Innovation

Table 5.5 Cumulative probability of stopping enrollment for futility

Efficacy Rate	Safety Rate	Sample Size (N)				
		125	150	200	250	225
0.64	**0.91**	0.003	0.007	0.011	0.018	0.034
0.66	**0.91**	0.005	0.009	0.010	0.012	0.022
0.68	**0.91**	0.003	0.004	0.006	0.009	0.013
0.70	**0.91**	0.003	0.003	0.006	0.009	0.012
0.72	**0.91**	0.003	0.004	0.012	0.015	0.020
0.74	**0.91**	0	0.003	0.006	0.011	0.019
0.64	**0.92**	0.002	0.004	0.008	0.013	0.019
0.66	**0.92**	0	0.004	0.005	0.009	0.012
0.68	**0.92**	0.001	0.001	0.003	0.003	0.005
0.70	**0.92**	0	0	0.001	0.002	0.002
0.72	**0.92**	0	0.003	0.004	0.005	0.006
0.74	**0.92**	0	0	0	0	0.001
0.64	**0.93**	0.004	0.006	0.009	0.012	0.020
0.66	**0.93**	0.003	0.003	0.003	0.005	0.006
0.68	**0.93**	0.001	0.001	0.001	0.001	0.001
0.70	**0.93**	0	0	0	0	0
0.72	**0.93**	0	0	0	0	0
0.74	**0.93**	0	0	0	0	0
0.64	**0.94**	0.004	0.007	0.010	0.011	0.018
0.66	**0.94**	0.001	0.002	0.003	0.004	0.005
0.68	**0.94**	0	0	0	0	0
0.70	**0.94**	0	0	0.001	0.001	0.001
0.72	**0.94**	0	0	0	0	0
0.74	**0.94**	0	0	0	0	0

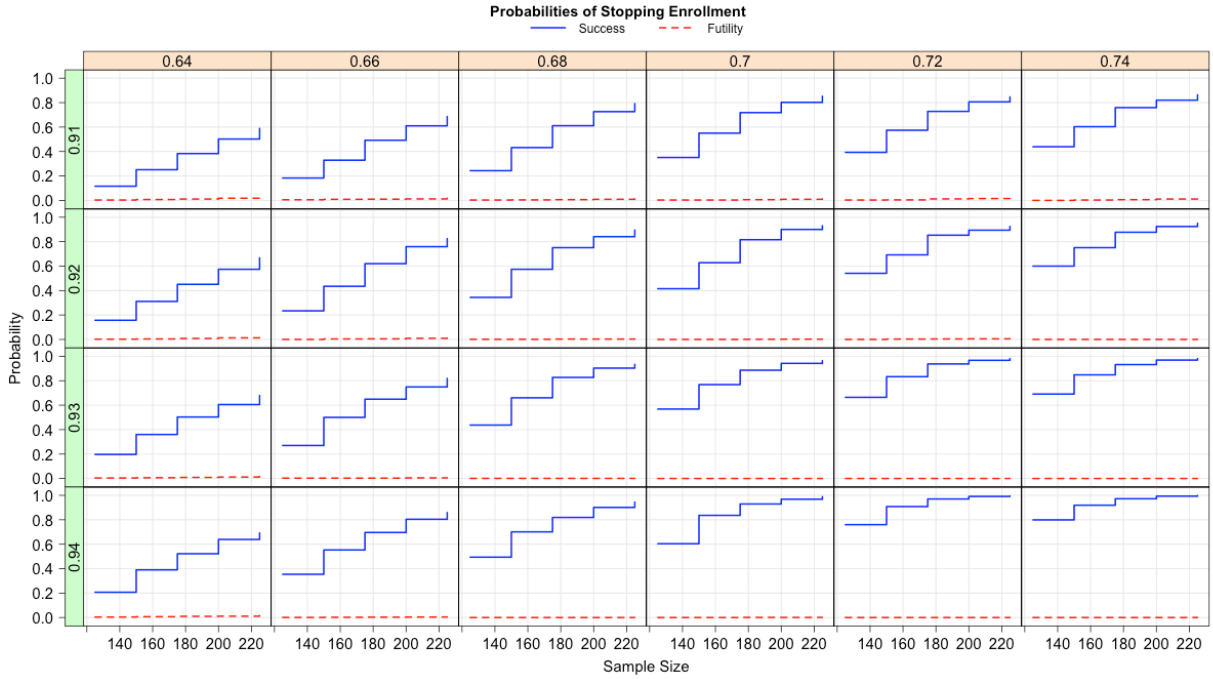

Figure XX: Proportion of trials that stopped at each interim sample size. Each column in the grid is specific to an efficacy scenario and each row is for a safety scenario. The possible sample sizes are displayed along the x-axis, and the height of the line shows the proportion of trials that stopped at or before that sample size. The blue line indicates trials that stopped for expected success and the red line indicates trials that stopped for futility.

Tables 6.5 and 5.7 contain the operating characteristics for the same set of scenarios with faster and slower accrual rates.

Berry Consultants
Statistical Innovation

Table 5.6 Positive Scenario Trial Outcomes (12 patients/month)

Efficacy Rate	Safety Rate	Mean N	Early Success	Late Success	Total Success	Early Futility	Late Failure
0.64	**0.91**	221.3	0.592	0.251	0.843	0.029	0.124
0.66	**0.91**	208.5	0.732	0.174	0.906	0.023	0.069
0.68	**0.91**	196.2	0.766	0.160	0.926	0.022	0.047
0.70	**0.91**	189.5	0.798	0.134	0.932	0.019	0.039
0.72	**0.91**	181.1	0.753	0.167	0.920	0.028	0.044
0.74	**0.91**	174.7	0.700	0.241	0.941	0.018	0.037
0.64	**0.92**	215.7	0.565	0.303	0.868	0.017	0.106
0.66	**0.92**	201.9	0.750	0.203	0.953	0.005	0.036
0.68	**0.92**	186.0	0.778	0.199	0.977	0.004	0.017
0.70	**0.92**	177.9	0.797	0.183	0.98	0.003	0.014
0.72	**0.92**	168.4	0.738	0.243	0.981	0.003	0.015
0.74	**0.92**	163.6	0.672	0.304	0.976	0.006	0.014
0.64	**0.93**	211.5	0.552	0.32	0.872	0.012	0.104
0.66	**0.93**	197.7	0.713	0.254	0.967	0.006	0.025
0.68	**0.93**	180.4	0.764	0.226	0.990	0.001	0.006
0.70	**0.93**	168.1	0.767	0.231	0.998	0	0.002
0.72	**0.93**	159.3	0.700	0.297	0.997	0	0.001
0.74	**0.93**	151.6	0.563	0.430	0.993	0.001	0.004
0.64	**0.94**	209.8	0.573	0.314	0.887	0.008	0.089
0.66	**0.94**	191.7	0.675	0.289	0.964	0.001	0.031
0.68	**0.94**	175.4	0.722	0.269	0.991	0.002	0.006
0.70	**0.94**	163.6	0.708	0.288	0.996	0	0.001
0.72	**0.94**	150.3	0.616	0.381	0.997	0	0
0.74	**0.94**	143.8	0.516	0.484	1	0	0

Berry Consultants
Statistical Innovation

Table 5.7 Positive Scenario Trial Outcomes (3 patients/month)

Efficacy Rate	Safety Rate	Mean N	Early Success	Late Success	Total Success	Early Futility	Late Failure
0.64	**0.91**	191.0	0.710	0.133	0.843	0.039	0.113
0.66	**0.91**	179.8	0.827	0.067	0.894	0.028	0.069
0.68	**0.91**	169.5	0.899	0.034	0.933	0.020	0.041
0.70	**0.91**	162.0	0.915	0.023	0.938	0.019	0.034
0.72	**0.91**	160.6	0.931	0.005	0.936	0.016	0.045
0.74	**0.91**	157.6	0.944	0.002	0.946	0.012	0.038
0.64	**0.92**	183.5	0.745	0.133	0.878	0.040	0.074
0.66	**0.92**	169.7	0.863	0.088	0.951	0.010	0.036
0.68	**0.92**	156.2	0.931	0.049	0.98	0.005	0.013
0.70	**0.92**	147.9	0.964	0.022	0.986	0.004	0.008
0.72	**0.92**	145.2	0.972	0.011	0.983	0.002	0.013
0.74	**0.92**	142.9	0.984	0	0.984	0.001	0.013
0.64	**0.93**	180.0	0.747	0.151	0.898	0.031	0.062
0.66	**0.93**	160.5	0.886	0.086	0.972	0.006	0.019
0.68	**0.93**	148.7	0.948	0.044	0.992	0.001	0.005
0.70	**0.93**	140.6	0.98	0.014	0.994	0.003	0
0.72	**0.93**	136.8	0.988	0.007	0.995	0.001	0.002
0.74	**0.93**	134.6	0.993	0.005	0.998	0.001	0.001
0.64	**0.94**	175.1	0.742	0.151	0.893	0.035	0.067
0.66	**0.94**	155.8	0.867	0.096	0.963	0.006	0.024
0.68	**0.94**	145.0	0.94	0.052	0.992	0.001	0.005
0.70	**0.94**	135.3	0.972	0.027	0.999	0	0
0.72	**0.94**	130.6	0.986	0.013	0.999	0	0
0.74	**0.94**	129.8	0.998	0.002	1	0	0

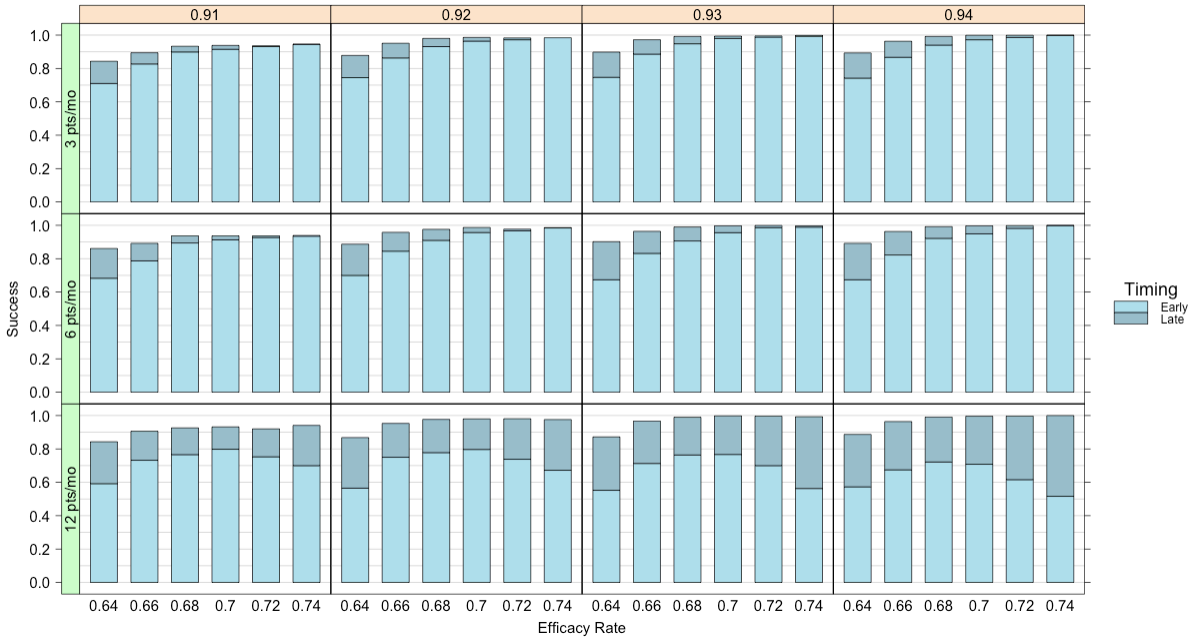

Figure XX: the proportion of successful trials for each scenario and each accrual rate. Each column in the grid is specific to a different safety rate, and each row is for an accrual rate. The efficacy rates are displayed along the x-axis. The height of each bar represents the power for the underlying rates of efficacy and safety. The light blue indicates the proportion of trials that claimed early success, and the dark blue represents trials that won late (after complete follow-up of all enrolled subjects).

We see that the probability of success increases within each panel as the efficacy rate increases, and that a larger proportion of these successful trials tend to be early. The exception to this trend is when the accrual rate is fast. Under a fast accrual rate, the trials with large effect sizes often stop enrollment at the first interim analysis, and therefore the minimum number of patients with complete follow-up is often only achieved after full follow-up of all subjects.

6.0 Control of Type I Error

In this section, we vary the scenarios and accrual rate assumptions in order to ensure appropriate control of type 1 error. In these null scenarios there is no underlying effect on one or both endpoints. We include cases that are deemed “worst case scenarios” for type I error inflation. The set of profiles is shown in Table 6.1.

In the first row, the efficacy and safety rates are exactly equal to their respective OPC’s. In scenarios 2 and 3, the efficacy rate is equal to the OPC and the safety rate is higher than the OPC. Scenarios 4 and 5 represent the opposite cases where the safety rate is equal to the OPC and the efficacy rate is higher.

Table 6.1 Null Scenario Profiles

Failure-free Rate	Hazard Rate (λ)			AE-free Rate	Hazard Rate (γ)		
	(0, 2]	(2, 8]	(8, 39]		(0, 1]	(1, 4]	(4, 26]
0.54	0.1998	0.0090	0.0052	0.84	0.0593	0.0297	0.0012
0.54	0.1998	0.0090	0.0052	0.92	0.0284	0.0142	0.0006
0.54	0.1998	0.0090	0.0052	0.999	0.0003	0.0002	0.0000
0.77	0.0847	0.0038	0.0022	0.84	0.0593	0.0297	0.0012
0.999	0.0003	0.0000	0.0000	0.84	0.0593	0.0297	0.0012

We simulate each of the null scenarios both with the futility rule enabled, and again with the futility rule disabled.

Each scenario was run for 5,000 simulated trials under different assumptions for the accrual rate.

6.1 Binding Futility Stopping Rule

Table 6.2 shows the operating characteristics with the futility rule enabled.

Because the trial requires success on both endpoints, type 1 error rates are very small when both endpoints are exactly equal to their respective OPCs. The type I error is below the 5% level in all scenarios.

Table 6.2 Null Scenario Outcomes (binding futility)

Accrual (pts/mo)	Efficacy Rate	Safety Rate	Mean N	Early Success	Late Success	Total Success	Early Futility	Late Failure
3	0.54	0.84	138.5	0.0006	0.0008	0.0014	0.9854	0.0132
	0.54	0.92	176.2	0.0110	0.0184	0.0294	0.8238	0.1464
	0.54	0.999	177.1	0.0134	0.0188	0.0322	0.8240	0.1422
	0.77	0.84	166.4	0.0458	0	0.0458	0.8672	0.0854
	0.999	0.84	164.7	0.0490	0	0.0490	0.8750	0.0754
6	0.54	0.84	140.9	0.0006	0.0012	0.0018	0.9822	0.0160
	0.54	0.92	186.3	0.0074	0.0196	0.0270	0.7704	0.2008
	0.54	0.999	186.7	0.0114	0.0196	0.0310	0.7634	0.2024
	0.77	0.84	168.7	0.0416	0.0000	0.0416	0.8574	0.0998
	0.999	0.84	169.9	0.0486	0.0004	0.0490	0.8514	0.0988
12	0.54	0.84	148.3	0	0.0016	0.0016	0.9650	0.0334
	0.54	0.92	205.9	0.0040	0.0242	0.0282	0.6106	0.3590
	0.54	0.999	208.7	0.0036	0.0224	0.0260	0.5944	0.3770
	0.77	0.84	172.4	0.0326	0.0064	0.0390	0.8492	0.1098
	0.999	0.84	172.6	0.0312	0.0104	0.0416	0.8456	0.1110

6.2 Nonbinding Futility Stopping Rule

Table 6.3 shows the operating characteristics when early futility is not enabled. Even when the trial does not stop early for futility, the type I error rate remains below the 5% level.

Table 6.3 Null Scenario Outcomes (nonbinding futility)

Accrual (pts/mo)	Efficacy Rate	Safety Rate	Mean N	Early Success	Late Success	Total Success	Early Futility	Late Failure
3	0.54	0.84	250.0	0.0002	0.0002	0.0004	0	0.9996
	0.54	0.92	248.5	0.0138	0.0188	0.0326	0	0.9658
	0.54	0.999	248.1	0.0160	0.0170	0.0330	0	0.9656
	0.77	0.84	246.8	0.0464	0	0.0464	0	0.9522
	0.999	0.84	246.5	0.0484	0	0.0484	0	0.9504
6	0.54	0.84	250.0	0	0.0008	0.0008	0	0.9992
	0.54	0.92	249.1	0.0078	0.0174	0.0252	0	0.9738
	0.54	0.999	248.6	0.0092	0.0196	0.0288	0	0.9692
	0.77	0.84	247.5	0.0402	0	0.0402	0	0.9572
	0.999	0.84	247.2	0.0430	0.0010	0.0440	0	0.9530
12	0.54	0.84	250.0	0	0.0006	0.0006	0	0.9994
	0.54	0.92	249.4	0.0040	0.0208	0.0248	0	0.9724
	0.54	0.999	249.3	0.0038	0.0236	0.0274	0	0.9696
	0.77	0.84	247.6	0.0358	0.0054	0.0412	0	0.9556
	0.999	0.84	248.2	0.0268	0.0072	0.0340	0	0.9626

7.0 Sensitivity to Model Misspecification

In this section, we present simulations in which the data was simulated from models other than the 3-piece exponential. The analysis model remains the same as described in Section 2.

We simulate all scenarios in this section assuming an accrual rate of 6 patients per month. Each scenario is run for 1000 trials.

7.1 Single-piece exponential

We simulate the time-to-failure and time-to-AE data from an exponential model with constant hazard. The hazard rates are calibrated to achieve the desired success rates.

Table 7.1 PE1 Profiles

Failure-free Rate	Hazard Rate (λ)	AE-free Rate	Hazard Rate (γ)
	(0, 39]		(0, 26]
0.64	0.0114	0.91	0.0036
0.74	0.0077	0.91	0.0036
0.64	0.0114	0.94	0.0024
0.74	0.0077	0.94	0.0024

Table 7.2 Sensitivity Scenario Outcomes (single-piece exponential)

Efficacy Rate	Safety Rate	Mean N	Early Success	Late Success	Total Success	Early Futility	Late Failure
0.64	0.91	202.5	0.509	0.372	0.881	0.022	0.088
0.74	0.91	166.1	0.91	0.049	0.959	0.018	0.019
0.64	0.94	178.2	0.53	0.415	0.945	0.004	0.035
0.74	0.94	131.2	0.962	0.038	1	0	0

7.2 Four-piece exponential

In this section, the data are simulated from a piecewise exponential with 4 segments. For the efficacy endpoint, we assume break points at 2 weeks, 8 weeks, and 26 weeks.

The profiles are constructed by taking advantage of the relationship:

$$\theta = \exp(-[2\,\lambda_1 + (8-2)\,\lambda_2 + (26-8)\,\lambda_3 + (39-26)\,\lambda_4]).$$

Furthermore, we assume the following pattern for the hazard rates:

$$h_1 = (30)\,h_4$$

$$h_2 = (5)\,h_4$$

$$h_3 = (2)\,h_4,$$

$$h_4 = (1)\,h_4.$$

The resulting profiles are shown in Table 7.3.

Table 7.3 Four-piece exponential profiles

Failure-free Rate	Hazard Rate (λ)				AE-free Rate	Hazard Rate (γ)			
	(0, 2]	(2, 8]	(8, 26]	(26, 39]		(0, 1]	(1, 4]	(4, 12]	(12, 26]
0.64	0.1280	0.0073	0.0055	0.0037	0.91	0.0215	0.0108	0.0043	0.0004
0.74	0.0864	0.0049	0.0037	0.0025	0.91	0.0215	0.0108	0.0043	0.0004
0.64	0.1280	0.0073	0.0055	0.0037	0.94	0.0141	0.0071	0.0028	0.0003
0.74	0.0864	0.0049	0.0037	0.0025	0.94	0.0141	0.0071	0.0028	0.0003

Table 7.4 Sensitivity Scenario Outcomes (four-piece exponential)

Efficacy Rate	Safety Rate	Mean N	Early Success	Late Success	Total Success	Early Futility	Late Failure
0.64	0.91	206.8	0.629	0.256	0.885	0.049	0.064
0.74	0.91	171.7	0.937	0.023	0.960	0.018	0.022
0.64	0.94	190.0	0.659	0.284	0.943	0.022	0.034
0.74	0.94	133.6	0.991	0.009	1	0	0

References

Wilber DJ, Pappone C, Neuzil P, et al. Comparison of Antiarrhythmic Drug Therapy and Radiofrequency Catheter Ablation in Patients with Paroxysmal Atrial Fibrillation: A Randomized Controlled Trial. *JAMA* 2010; 303(4): 333-340.

Appendix 1: Computational Details

The simulations were run using R version 2.15. Kaplan-Meier curves were fitted using the survival package version 2.37.4