Sample Size Calculation Approaches for a Group Sequential Clinical Trial

Sam VanSchalkwyk

January 22, 2018

1 Methods

The use of intravenous antibiotic delivery (IVAD) is predicted to reduce the rate of surgical site infection (SSI) compared with the currently used tumescent anesthesia antibiotic delivery (TAAD), in particular for patients receiving emergency gastrointestinal surgery. Let p_C be the proportion of patients receiving TAAD that get a SSI, and p_T the corresponding proportion for patients that received IVAD. Consider a one sided test, where

$$H_0 : p_C = p_T$$

and

$$H_1 : p_C > p_T.$$

Using a 2-stage group sequential test, we wish to control the Type I Error at $\alpha = 0.05$ with power of $1-\beta = 0.8$ at $p_C - p_T = \theta$ for various values of θ. Let Z_k be the test statistic calculated at stage k, where $k = 1, 2$. For information levels $\{I_1, I_2\}$, the test statistics (Z_1, Z_2) are approximately bivariate normal with $E(Z_k) = \theta\sqrt{I_k}$, $k = 1, 2$ and $Cov(Z_1, Z_2) = \sqrt{I_1/I_2}$. Then after collecting the required number of samples in the first stage, the test is

if $Z_1 \geq b_1$	stop, reject H_0
if $Z_1 \leq a_1$	stop, accept H_1
otherwise	continue to stage 2

and if the test continues to stage 2 then

if $Z_2 \geq b_2$	stop, reject H_0
if $Z_2 < a_2$	stop, accept H_1

where $a_2 = b_2$ so that the test ends by stage 2.

Determining the threshold values will change the arm size m for the test, which is the number of samples required from each group at each stage. For example, if $m = 100$, then 100 subjects are needed from the control group and another 100 subjects are needed for the treatment group at each stage. For the power family of tests, equally spaced increments of information are used, so we have $I_k = \frac{k}{2} I_2$, $k = 1, 2$. The power family is indexed by parameter Δ, and critical values are

$$\begin{aligned} b_k &= \tilde{C}_1(K, \alpha, \beta, \Delta)(k/K)^{\Delta - 0.5} \\ a_k &= \delta\sqrt{I_k} - \tilde{C}_2(K, \alpha, \beta, \Delta)(k/K)^{\Delta - 0.5} \end{aligned}$$

for $k = 1, 2$, where $K = 2$ is the number of trials in the group sequential test and δ is the reference value for $p_C - p_T$, or in other words the expected difference between the proportion of SSIs in the control and treatment groups. To ensure that $a_2 = b_2$, we must have that

$$I_2 = \frac{[\tilde{C}_1(K, \alpha, \beta, \Delta) + \tilde{C}_2(K, \alpha, \beta, \Delta)]^2}{\delta^2}$$

and values of $\tilde{C}_1(K, \alpha, \beta, \Delta)$ and $\tilde{C}_2(K, \alpha, \beta, \Delta)$ are chosen to satisfy the desired Type I error and power for the test.

Case	Risk	Effect Size	p_C	p_T
1	14%	50%	0.14	0.07
2	25%	50%	0.25	0.125
3	50%	50%	0.50	0.25
4	30%	50%	0.30	0.15
5	30%	40%	0.30	0.18

Table 1: Case Risk and Effect Sizes

	Case	OBF	Pocock	$\frac{\alpha}{2}, \frac{\beta}{2}$	$\frac{\alpha}{4}, \frac{\beta}{4}$
m	1	126	155	135	123
	2	64	78	68	62
	3	24	29	25	23
	4	50	62	54	49
	5	83	102	88	81
ESS H_1	1	401	392	384	397
	2	202	197	193	199
	3	75	73	71	74
	4	159	156	153	157
	5	263	257	252	260

Table 2: Arm Sizes and Expected Sample Sizes under H_1 (ESS H_1)

1.1 O'Brien-Fleming

The O'Brien-Fleming (OBF) test belongs to the power family with $\Delta = 0$.

1.2 Pocock

The Pocock test belongs to the power family with $\Delta = 0.5$. Compared with the O'Brien-Fleming test, the Pocock test has lower expected sample size under the alternative hypothesis, but a larger maximum sample size if the test does continue to the second stage.

1.3 Error Spending

The Error Spending approach provides the flexibility of choosing how much Type I and Type II error to spend at each stage. For this trial with 2 stages, we consider two possibilities of error spending. In one situation, we allow $\alpha_1 = \alpha/2$ and $\beta_1 = \beta/2$ to be spent at stage 1, with the remaining error spent at the second stage. Next we allow $\alpha_1 = \alpha/4$ and $\beta_1 = \beta/4$ to be spent at stage 1, and again the remaining error would be allocated to the second stage. With the error spending approach, it can be seen that arm sizes are not very large, and in particular with the appropriate choice of error spending, the expected sample size under the alternative hypothesis will be lower than that of the O'Brien-Fleming or Pocock tests.

2 Cases

There are 5 cases that will be considered for sample size calculation. For each case, let the risk be the rate at which patients receiving TAAD will have SSI. Let the effect size be the expected reduction of the risk of SSI for patients receiving IVAD compared to those receiving TAAD. Details on the risk and effect sizes for these 5 cases are displayed in Table 1. In the first 4 cases, assume a 50% effect size. That is, the hypothesized proportion of SSIs in the treatment group receiving IVAD is 50% of the corresponding hypothesized proportion in the control group receiving TAD. Assume a risk of 14%, 25%, 50%, and 30% for cases 1, 2, 3, and 4, respectively. For case 5, let the effect size be 40% with a 30% risk. Refer to Table 2 for arm sizes and expected sample sizes under H_1 for each of these 5 cases.