Pubh 8482: Sequential Analysis

Joseph S. Koopmeiners

Division of Biostatistics
University of Minnesota

Week 5

Course Summary

So far, we have discussed

- Group sequential procedures for two-sided tests
- Group sequential procedures for one-sided tests
- Group sequential procedures for two-sided tests with an inner-wedge

Course Summary

So far, we have discussed

- Group sequential procedures for two-sided tests
- Group sequential procedures for one-sided tests
- Group sequential procedures for two-sided tests with an inner-wedge

Course Summary

Some patterns have emerged

- We specify the general shape of the stopping boundaries and solve to identify boundaries that achieve the desired type-I and type-II error rates
- The maximum sample size must be increased relative to the fixed-sample size in order to achieve the desired power
- There is a trade-off between expected sample size and maximum sample size

Limitation

- To this point, we have assumed that interim analyses are equally spaced throughout the study
- This implicitly assumes that information accrues uniformly throughout the study
- Similar approaches can be used to develop stopping boundaries if the information available at each interim analysis is known a priori

Information accrual in practice

- In practice, it is often the case that the final information is known but the information available at the interim analyses is not
  - If subject accrual is uneven throughout the study
  - Information is proportional to the number of events in a survival analysis
- How do we develop stopping boundaries when the amount of information at the interim analyses is unknown?

The Error Spending Approach

- Slud and Wei (1982) introduced an error spending approach that guarantees the desired type-I error rate
  - Consider a two-sided test of $\delta = \mu_x - \mu_y$ in the setting of a two arm trial of two normally distributed random variables with known variance
  - Test the null hypothesis $H_0 : \delta = 0$
  - Assume that we will have K analyses
  - Let α be the overall type-I error rate

The Error Spending Approach: Stopping Boundaries

- Define critical values c_k for $k = 1, \ldots, K$
- For $k = 1, \ldots, K - 1$
  - If $|Z_k| > c_k$, stop and reject H_0
  - otherwise, continue to group $k + 1$
- For $k = K$
  - If $|Z_K| > c_K$, stop and reject H_0
  - otherwise, stop and fail to reject H_0

The Error Spending Approach: Critical Values

- The error spending approach partitions α into probabilities $\pi_1, \ldots, \pi_K$ that sum to α
- Let $I_1, \ldots, I_K$ be a general sequence of information
  - The critical values $c_1, \ldots, c_K$ are defined such that

    $$P\left(|Z_1| > c_1 | \delta = 0, I_1\right) = \pi_1$$

    and

    $$P\left(|Z_1| < c_1, \ldots, |Z_{k-1}| < c_{k-1}, |Z_k| > c_k | \delta = 0, I_1, \ldots, I_k\right) = \pi_k$$

    for $k = 2, \ldots, K$

Error Spending Example

- Consider a group sequential design with:
  - $\alpha = 0.05$
  - $K = 5$
  - Fixed-sample size of $n = 100$/group
  - 80% power

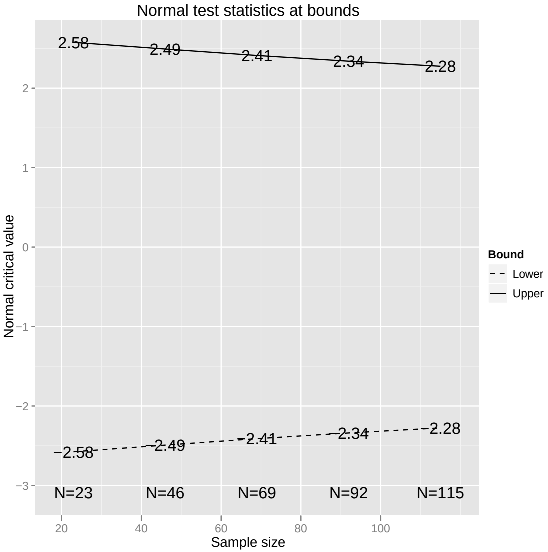
Error Spending Example 1

- Consider the following sequence of probabilities
  - $\pi_1 = 0.01$
  - $\pi_2 = 0.01$
  - $\pi_3 = 0.01$
  - $\pi_4 = 0.01$
  - $\pi_5 = 0.01$
- Assume that stopping times are evenly spaced throughout the study (i.e. after $n = 20$/group, $n = 40$/group, etc.)

Error Spending Example 1: Critical Values

- This results in the following critical values
  - $c_1 = 2.58$
  - $c_2 = 2.49$
  - $c_3 = 2.41$
  - $c_4 = 2.34$
  - $c_5 = 2.28$
- The maximum sample size must be inflated to $n = 115$/group to achieve 80% power

Error Spending Example 1: Critical Values



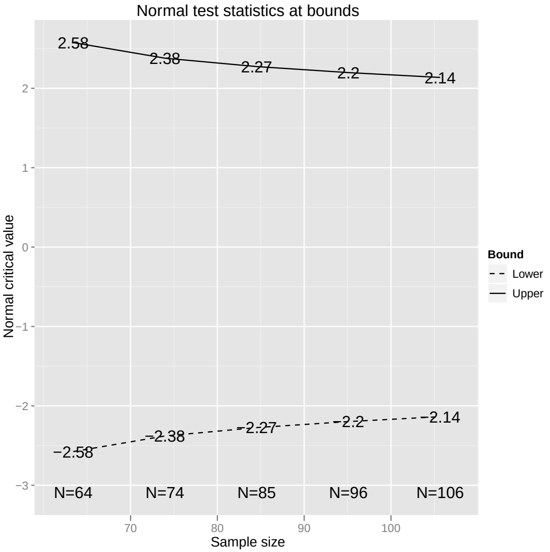
Error Spending Example 2

- Consider the same sequence of probabilities as before
- Assume that stopping times are evenly spaced throughout the second half of the study (i.e. after $n = 60$/group, $n = 70$/group, etc.)

Error Spending Example 2: Critical Values

- This results in the following critical values
  - $c_1 = 2.58$
  - $c_2 = 2.38$
  - $c_3 = 2.27$
  - $c_4 = 2.20$
  - $c_5 = 2.14$
- The maximum sample size must be inflated to $n = 106$/group to achieve 80% power

Error Spending Example 2: Critical Values



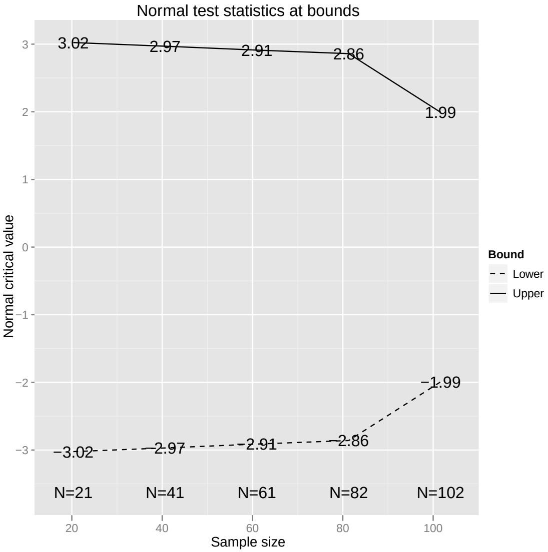
Error Spending Example 3

- Finally, consider the situation were a small amount of error is spent at the first interim analyses and a large amount at the final analysis
  - $\pi_1 = 0.0025$
  - $\pi_2 = 0.0025$
  - $\pi_3 = 0.0025$
  - $\pi_4 = 0.0025$
  - $\pi_5 = 0.04$
- Assume that stopping times are evenly spaced throughout the study (i.e. after $n = 20$/group, $n = 40$/group, etc.)

Error Spending Example 2: Critical Values

- This results in the following critical values
  - $c_1 = 3.02$
  - $c_2 = 2.97$
  - $c_3 = 2.91$
  - $c_4 = 2.86$
  - $c_5 = 1.99$
- The maximum sample size must be inflated to $n = 102$/group to achieve 80% power

Error Spending Example 3: Critical Values



Properties of Slud and Wei's approach

- It is clear that the overall type-1 error rate is α
- The first critical value is $c_1 = \Phi^{-1}\left(1 - \frac{\pi_1}{2}\right)$
- Additional critical values are found numerically
- A critical value c_k depends on the information available at the first k analyses but not the unobserved information $I_{k+1}, \ldots, I_K$

Limitations of Slud and Wei's approach

- The number of analyses, K, must be fixed in advance
- We might want flexibility in the amount of error spent at each interim analysis depending on the amount of information accrued since the previous analysis

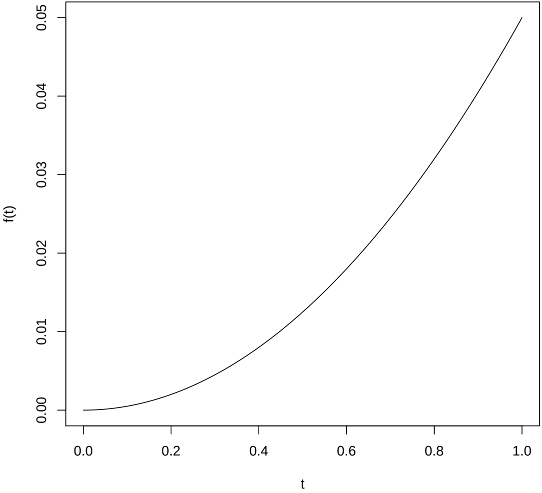
An alternate approach

- An alternate approach is the maximum information trial proposed by Lan and DeMets (1983)
- In this case, subjects are enrolled until a maximum information level is reached
- Error is spent according to an error spending function

Error Spending Function

- In the maximum information trial approach, information is partitioned using an error spending function
- Error spending functions have the following properties
  - non-decreasing
  - $f(0) = 0$
  - $f(1) = \alpha$, where α is the desired type 1 error rate

Error Spending Function: Example



Error Spending Function: Calculating critical values

- Let I_{max} be the target maximum information level
- Let $I_1, I_2, \ldots, I_k$ be a sequence of information values for the first k stopping times
- The error spending function is first translated into probabilities as in Slud and Wei's method
  - $\pi_1 = f\left(I_1/I_{max}\right)$
  - $\pi_k = f\left(I_k/I_{max}\right) - f\left(I_{k-1}/I_{max}\right)$ for $k = 2, 3, \ldots$
- $\pi_1, \pi_2, \ldots$ are translated into $c_1, c_2, \ldots$ as in Slud and Wei's method

Error Spending Function: Critical values example

- Consider the simple error spending function $f(t) = \alpha t$ with $\alpha = 0.05$
- Recall for our two sample case, that $I_k = \frac{2 * n_k}{\sigma^2}$
- Let the target maximum information level be $I_{max} = \frac{2 * 100}{\sigma^2}$

Error Spending Function: Critical values example

- Let's assume that the first interim analysis is at $n = 20$/group
  - $\pi_1 = f\left(I_1/I_{max} = .2\right) = 0.01$
  - This corresponds to a critical value of $c_1 = \Phi\left(1 - 0.01/2\right)^{-1} = 2.58$
- Let's assume the second interim analysis is at $n = 50$/group
  - $\pi_2 = f\left(I_2/I_{max}\right) - f\left(I_1/I_{max}\right) = f\left(.5\right) - f\left(.2\right) = 0.015$
  - This corresponds to a critical value of $c_2 = 2.38$
- We'll consider two scenarios for the remainder of the trial

Error Spending Function: Scenario 1

- No additional interim analyses before study completion
  - $\pi_3 = f(I_3/I_{max}) - f(I_2/I_{max}) = f(1) - f(.5) = 0.025$
  - This corresponds to a critical value of $c_3 = 2.14$
- What if there is an additional interim analysis?

Error Spending Function: Scenario 2

- Add an additional interim analysis at $n = 75$/group
  - $\pi_3 = f(I_3/I_{max}) - f(I_2/I_{max}) = f(.75) - f(.5) = 0.0125$
  - This corresponds to a critical value of $c_3 = 2.32$
- Final analysis at study completion
  - $\pi_4 = f(I_4/I_{max}) - f(I_3/I_{max}) = f(1) - f(.75) = 0.0125$
  - This corresponds to a critical value of $c_4 = 2.24$
- The critical value is larger than before $(2.24 vs. 2.14)$
- The additional interim analysis spent an additional amount of type-1 error leaving less available for the final analysis

Properties of Error Spending Functions

- We need not specifying the number or timing of interim analyses in advance
- Critical values depend on the number of previous interim analyses
- Critical values depend on the sequence of information available at previous interim analyses
- Critical values do not depend on the number of interim analyses or sequence of information for the remainder of the trial

Properties of Error Spending Functions

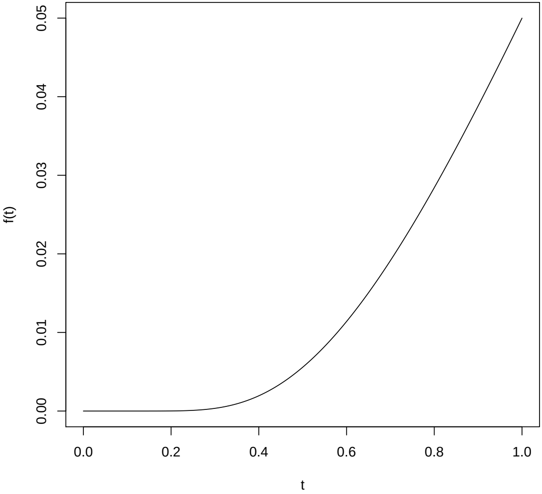
- We need not specifying the number or timing of interim analyses in advance
- Critical values depend on the number of previous interim analyses
- Critical values depend on the sequence of information available at previous interim analyses
- Critical values do not depend on the number of interim analyses or sequence of information for the remainder of the trial

Error Spending Function: Example 1

Lan and DeMets (1983) show that the following error spending function results in critical values similar to the O'Brien-Fleming boundaries

$$f(t) = min\left[2 - 2\Phi\left(Z_{1-\alpha/2}/\sqrt{t}\right), \alpha\right]$$

Error Spending Function: Example 1

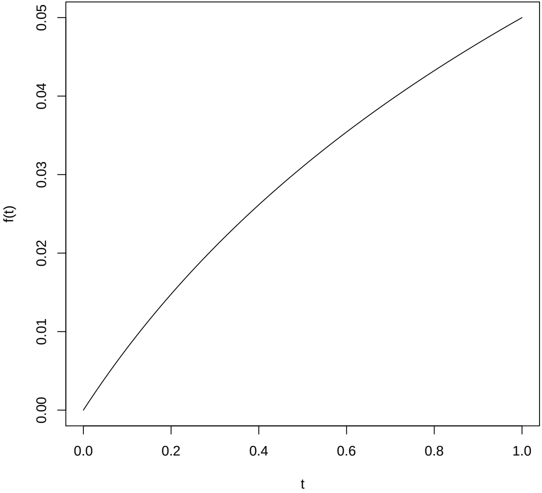


Error Spending Function: Example 2

Lan and DeMets (1983) also show that the following error spending function results in critical values similar to the Pocock boundaries

$$f(t) = min\left[\alpha log\left(1 + (e - 1)\, t\right), \alpha\right]$$

Error Spending Function: Example 2

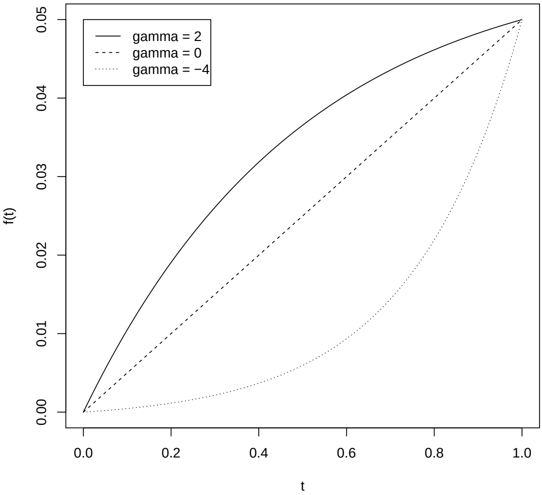


Error Spending Function: Example 3

Hwang, Shih and DeCani (1990) introduce a family of error spending functions indexed by a parameter γ

$$f(t) = \begin{cases} \alpha \left(1 - e^{-\gamma t}\right)\left(1 - e^{-\gamma}\right) & \text{if } \gamma \neq 0 \\ \alpha t & \text{if } \gamma = 0. \end{cases}$$

Error Spending Function: Example 3



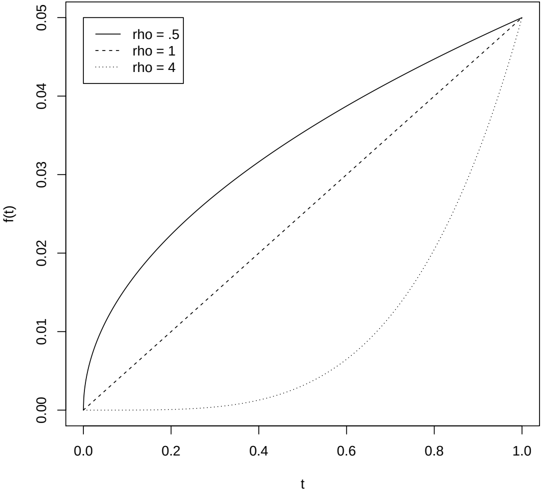
Error Spending Function: Example 4

Kim and DeMets (1987) present the following error spending function

$$f(t) = \alpha t^{\rho}$$

for $\rho > 0$

Error Spending Function: Example 4



Stopping Boundaries for Error Spending Function

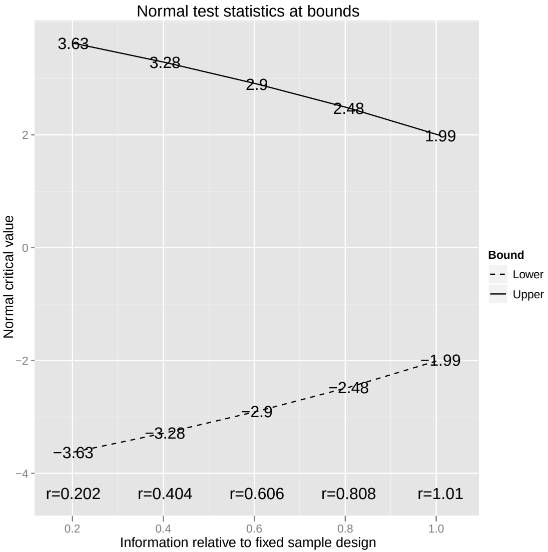
- We will focus on the error spending functions proposed by Hwang-Shih and Decani and Kim and DeMets
- We will consider a two-sided test with $\alpha = 0.05$
- While the number of interim analyses need not be specified when using the error spending approach, we will consider designs with $K = 5$ to illustrate the shape of the boundaries

Hwang, Shih and DeCani: Equally spaced stopping times

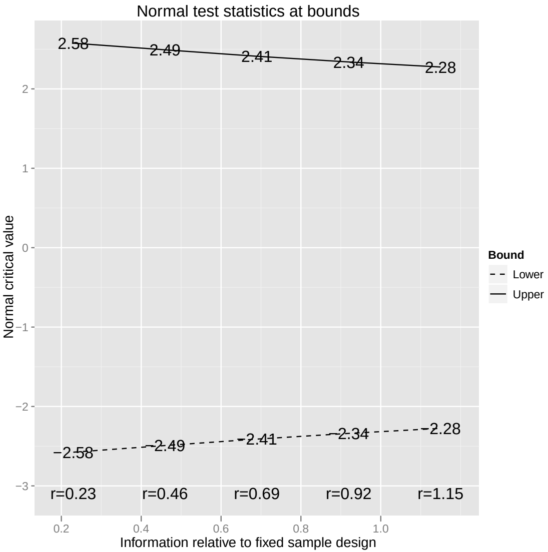
- First consider the Hwang, Shih and DeCani error spending function
- Consider 5, equally spaced interim analyses

k	$\gamma = -6$	$\gamma = 0$	$\gamma = 10$
1	3.63	2.58	2.02
2	3.28	2.49	2.53
3	2.90	2.41	3.01
4	2.48	2.34	3.47
5	1.99	2.28	3.90

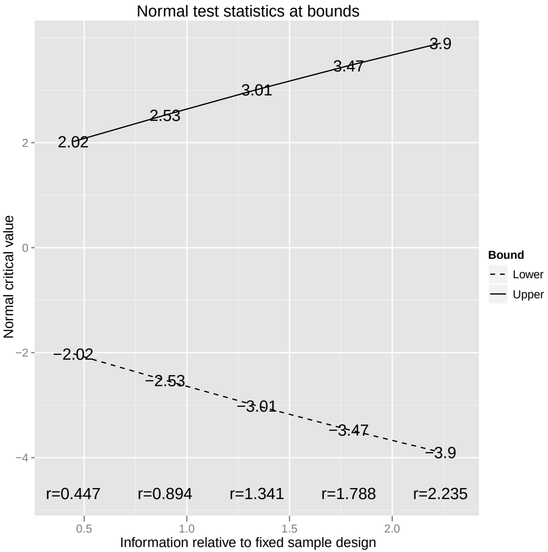
Equally-spaced Hwang, Shih and DeCanni Boundaries with $\gamma = -6$



Equally-spaced Hwang, Shih and DeCanni Boundaries with $\gamma = 0$



Equally-spaced Hwang, Shih and DeCanni Boundaries with $\gamma = 10$

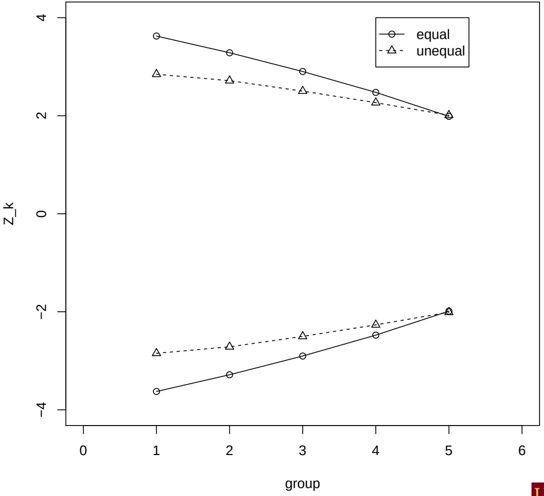


Hwang, Shih and DeCani: Unequally spaced stopping times

- What if interim analyses are clustered towards the end of the trial
  - Interim analyses at $n = 60, 70, 80, 90 and 100$

k	equal	unequal
1	3.63	2.85
2	3.28	2.71
3	2.90	2.50
4	2.48	2.27
5	1.99	2.01

Comparing equal and unequally spaced interim analyses



Hwang, Shih and DeCanni spending function: summary

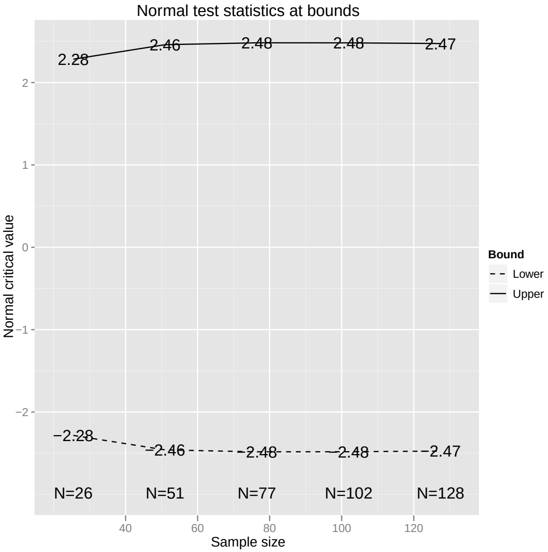
- Increasing γ implies that more error is spent early in the trial and less is available late in the trial
- Larger values of γ lead to smaller critical values early on but larger critical values late in the trial
- Equally spaced increments spend small amount of error early, saving more for later in the trial

Kim and DeMets: Equally spaced stopping times

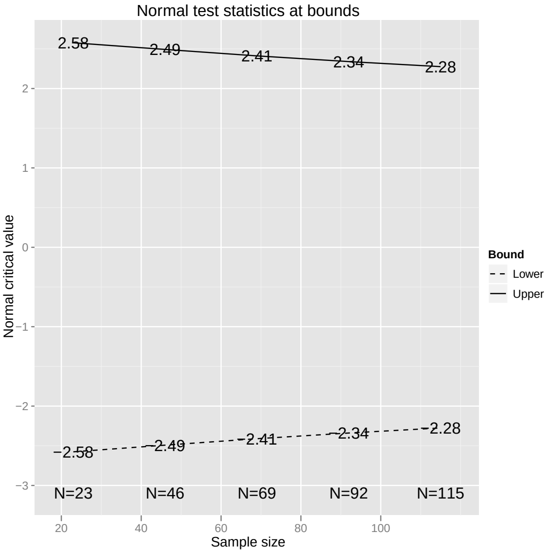
- First consider the Kim and DeMets error spending function
- Consider 5, equally spaced interim analyses

k	$\rho = 0.5$	$\rho = 1$	$\rho = 4$
1	2.28	2.58	3.94
2	2.46	2.49	3.23
3	2.48	2.41	2.75
4	2.48	2.34	2.36
5	2.47	2.28	2.01

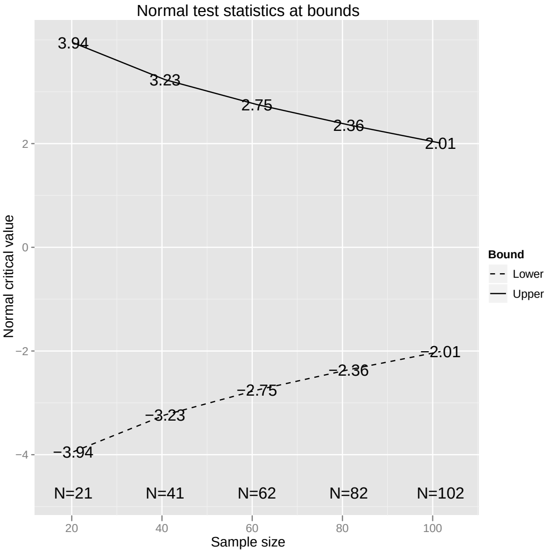
Equally-spaced Kim and DeMets Boundaries with $\rho = 0.5$



Equally-spaced Kim and DeMets Boundaries with $\rho = 1$



Equally-spaced Kim and DeMets Boundaries with $\rho = 10$

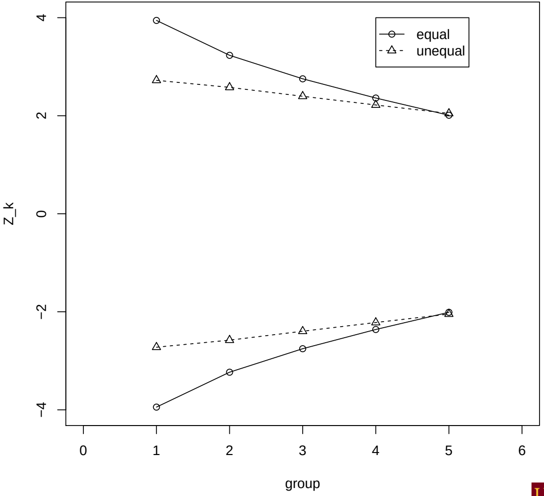


Kim and DeMetsi: Unequally spaced stopping times

- What if interim analyses are clustered towards the end of the trial
  - Interim analyses at $n = 60, 70, 80, 90 and 100$

k	equal	unequal
1	3.94	2.72
2	3.23	2.58
3	2.75	2.40
4	2.36	2.22
5	2.01	2.05

Comparing equal and unequally spaced interim analyses



Kim and DeMets spending function: summary

- Increasing ρ implies that less error is spent early in the trial and more is available late in the trial
- Smaller values of ρ lead to smaller critical values early on but larger critical values late in the trial
- Equally spaced increments spend small amount of error early, saving more for later in the trial

Error Spending Functions and Power

- We have seen that using a groups sequential stopping rule results in decreased power compared to a fixed-sample design
- We must increase the maximum sample size of a clinical trial in order to achieve the same power as a fixed-sample design
- The power of a group sequential design will depend on the number and timing of the interim analyses
- How should we proceed when the point of error spending functions is that we need not specify the number and timing of interim analyses?

Error Spending Functions and Power

- In practice, the number and timing of interim analyses is specified in advance for a clinical trial
  - DSMB meetings scheduled yearly or every six months for length of trial
- What is unknown is the information available at each interim analysis?
- The simplest approach is to assume uniform information growth when design the study

Error Spending Functions and Power: Example

- Consider a two-sided test monitored using an error-spending function
  - $\alpha = 0.05$
  - We will power the study to detect an effect size of $0.2 * \sigma$
  - We would need a sample size of 393/group to achieve 80% power
  - We would need a sample size of 526/group to achieve 90% power
- The sample size inflation factor will depend on the error spending function and parameter values

Sample Size Inflation Factor: Hwang, Shih and DeCanni

Hwang, Shih and DeCanni spending function with 80% power

K	$\gamma = -3$	$\gamma = 0$	$\gamma = 3$
2	1.017	1.082	1.233
3	1.028	1.117	1.32
4	1.036	1.137	1.366
5	1.041	1.15	1.394
8	1.05	1.17	1.436
10	1.054	1.178	1.45

Sample Size Inflation Factor: Hwang, Shih and DeCanni

Kim and DeMets spending function with 90% power

K	$\gamma = -3$	$\gamma = 0$	$\gamma = 3$
2	1.016	1.075	1.211
3	1.026	1.107	1.289
4	1.033	1.124	1.329
5	1.038	1.136	1.354
8	1.046	1.155	1.392
10	1.05	1.162	1.405

Sample Size Inflation Factor: Kim and DeMets

Kim and DeMets spending function with 80% power

K	$\rho = .5$	$\rho = 1$	$\rho = 3$
2	1.162	1.082	1.01
3	1.222	1.117	1.02
4	1.254	1.137	1.027
5	1.274	1.15	1.032
8	1.306	1.17	1.041
10	1.317	1.178	1.045

Sample Size Inflation Factor: Kim and DeMets

Kim and DeMets spending function with 90% power

K	$\rho = .5$	$\rho = 1$	$\rho = 3$
	1.146	1.075	1.009
	1.2	1.107	1.018
	1.229	1.124	1.025
	1.247	1.136	1.03
	1.275	1.155	1.039
	1.285	1.162	1.042

What if the interim analyses are not equally spaced?

- The above sample size inflation factors assume equally spaced interim analyses
- What if the interim analyses are not evenly spaced?
- Assume, instead, that interim analyses are planned at approximately:
    - 60% of the information
    - 70% of the information
    - 80% of the information
    - 90% of the information
    - 100% of the information

Unevenly Spaced Interim Analyses: Hwang, Shih and DeCanni

Consider stopping boundaries developed using the Hwang, Shih and DeCanni error spending function with K = 5 and 80% power

γ	Even	Uneven
-3	1.041	1.045
0	1.15	1.136
3	1.394	1.304

Unevenly Spaced Interim Analyses: Kim and DeMets

Consider stopping boundaries developed using the Kim and DeMets error spending functions with K = 5 and 80% power

γ	Even	Uneven
0.5	1.274	1.212
1	1.15	1.136
3	1.032	1.042

Sample Size Inflation for Error Spending Functions: Summary

- Sample size inflation increased as γ increased for the Hwang, Shih and DeCanni Error spending function and as ρ decreased for the Kim and DeMets error spending function
- More generally, a large sample size inflation is required when more error is spent early in the trial
- The unevenly spaced stopping times required a smaller sample size inflation when more error was spent early in the trial and a larger inflation when more error was available at the end of the trial

Sample Size Inflation for Error Spending Functions: A final thought

- Initially, we discussed planning the study assuming equally spaced interim analysis
- We then noted that changing the timing of the interim analyses would change the power/sample size inflation factor
- You have to make assumptions about the timing of interim analysis (whether equal or not) to identify the correct sample size
- We'll see you the effect of being wrong on the next homework
- Remember, though, that the type-I error rate will be controlled regardless of the timing of interim analyses

Expected Sample Size for Error Spending Functions

- As with standard group sequential designs, we'll evaluate the savings due to an error spending function by considering the expected sample size
- The expected sample size will depend on several parameters
  - α
  - β
  - K
  - Error spending function and parameter
  - True difference between groups

Expected Sample Size for Error Spending Function

Expected sample size as percentage of fixed sample size for Hwang, Shih and DeCanni error spending function for varies values of γ with

- $\alpha = 0.05$
- K = 5
- equally spaced interim analyses
- 80% power to detect an effect size of δ

γ	$0 * \delta$	$.5 * \delta$	δ	$1.5 * \delta$
-3	103.1	98.3	79.7	56.3
0	112.7	104.8	78.5	51
3	135.1	122.8	84.1	50.2

Expected Sample Size for Error Spending Function

Expected sample size as percentage of fixed sample size for Hwang, Shih and DeCanni error spending function for varies values of K with

- $\alpha = 0.05$
- $\gamma = -3$
- equally spaced interim analyses
- 80% power to detect an effect size of δ

γ	$0 * \delta$	$.5 * \delta$	δ	$1.5 * \delta$
2	101.3	99	87.9	68.6
3	102.1	98.5	83.3	61.6
4	102.7	98.4	81.1	58.2
5	103.1	98.3	79.7	56.3
8	103.9	98.3	77.8	53.4
10	104.2	98.3	77.2	52.5

Expected Sample Size for Error Spending Function

Expected sample size as percentage of fixed sample size for Hwang, Shih and DeCanni error spending function for varies values of K with

- $\alpha = 0.05$
- $\gamma = 3$
- equally spaced interim analyses
- 80% power to detect an effect size of δ

γ	$0 * \delta$	$.5 * \delta$	δ	$1.5 * \delta$
2	120.8	112.6	88.7	68.1
3	128.6	118	85.8	57.3
4	132.6	121	84.6	52.6
5	135.1	122.8	84.1	50.2
8	138.9	125.7	83.4	47
10	140.2	126.7	83.2	46

Expected Sample Size for Error Spending Function

Expected sample size as percentage of fixed sample size for Kim and DeMets error spending function for varies values of ρ with

- $\alpha = 0.05$
- K = 5
- equally spaced interim analyses
- 80% power to detect an effect size of δ

ρ	$0 * \delta$	$.5 * \delta$	δ	$1.5 * \delta$
0.5	123.9	114.1	81.6	49.9
1	112.7	104.8	78.5	51
3	102.4	97.9	80.6	58.8

Expected Sample Size for Error Spending Function

Expected sample size as percentage of fixed sample size for Kim and DeMets error spending function for varies values of K with

- $\alpha = 0.05$
- $\rho = 0.5$
- equally spaced interim analyses
- 80% power to detect an effect size of δ

γ	$0 * \delta$	$.5 * \delta$	δ	$1.5 * \delta$
2	114.1	107.4	86.4	66
3	119.4	110.8	83.3	56
4	122.2	112.8	82.2	52
5	123.9	114.1	81.6	49.9
8	126.7	116.2	81.1	47.4
10	127.7	117	81	46.7

Expected Sample Size for Error Spending Function

Expected sample size as percentage of fixed sample size for Kim and DeMets error spending function for varies values of K with

- $\alpha = 0.05$
- $\rho = 3$
- equally spaced interim analyses
- 80% power to detect an effect size of δ

γ	$0 * \delta$	$.5 * \delta$	δ	$1.5 * \delta$
2	100.7	98.9	89.5	70.7
3	101.4	98.3	84.6	64.7
4	102	98	82.1	61
5	102.4	97.9	80.6	58.8
8	103.1	97.9	78.5	55.6
10	103.4	97.9	77.8	54.6

Expected Sample Size for Error Spending Function: Summary

- Error spending functions that spend less error early have a smaller expected sample size with smaller differences but a larger expected sample size if the difference is large
  - The difference under the alternative is modest compared to the difference under the null

Expected Sample Size for Error Spending Function: Summary

- For error spending functions that spend a large amount of error early
  - Expected sample size increases with K for small or no difference
  - Expected sample size decreases with K for larger differences

Expected Sample Size for Error Spending Function: Summary

- For error spending functions that spend a smaller amount of error early
  - Expected sample size increases with K only under the null
  - Expected sample size decreases with K otherwise

Sequential monitoring of SBP using an error spending function

Consider sequential monitoring of SBP at 24 months in the Mr Fit study using the error spending approach

- $\alpha = 0.05$
- $n = 400$/group
- Assume that σ is known and equal to 14
- We will use the Kim and DeMets error spending function with $\rho = 3$

$$f(t) = \alpha t^3$$

- We need not specify the timing of interim analyses in advance

Sequential monitoring of SBP: first interim analysis

Let's assume that the first interim analysis occurs with 80 subjects per group

- $I_1/I_{max} = 80/400 = .2$
- $f(.2) = 0.0004$
- This corresponds to a critical value of $c_1 = 3.54$
- $Z_1 = .875$
- Continue the trial

Sequential monitoring of SBP: second interim analysis

The second interim analysis occurs with 140 subjects per group

- $I_2/I_{max} = 140/400 = .35$
- $f(.35) = 0.0021$
- $f(.35) - f(.2) = 0.0017$
- This corresponds to a critical value of $c_2 = 3.11$
- $Z_2 = 2.86$
- Continue the trial

Sequential monitoring of SBP: third interim analysis

The second interim analysis occurs with 280 subjects per group

- $I_3/I_{max} = 280/400 = .70$
- $f(.70) = 0.0150$
- $f(.70) - f(.35) = 0.0017$
- This corresponds to a critical value of $c_3 = 2.41$
- $Z_3 = 5.82$
- Stop and reject the null hypothesis

Sequential monitoring of SBP: Summary

- The trial stops at the third interim analysis
- We reject the null hypothesis and conclude that subjects in the control group have a significantly higher SBP than subjects in the experimental condition
- We use a sample size of 280 subjects per group, which is a savings of 120 subjects per group compared to the fixed-sample design

Sequential monitoring of DBP using an error spending function

Consider sequential monitoring of DBP at 24 months in the Mr Fit study using the error spending approach

- $\alpha = 0.05$
- $n = 400$/group
- Assume that σ is known and equal to 8
- We will use the Kim and DeMets error spending function with $\rho = .9$

$$f(t) = \alpha t^{.9}$$

- This will spend more error early in the trial than when $\rho = 3$

Sequential monitoring of DBP: first interim analysis

Let's assume that the first interim analysis occurs with 40 subjects per group

- $I_1/I_{max} = 40/400 = .1$
- $f(.1) = 0.0063$
- This corresponds to a critical value of $c_1 = 2.73$
- $Z_1 = 1.44$
- Continue the trial

Sequential monitoring of SBP: second interim analysis

The second interim analysis occurs with 160 subjects per group

- $I_2/I_{max} = 160/400 = .40$
- $f(.40) = 0.0219$
- $f(.40) - f(.1) = 0.0156$
- This corresponds to a critical value of $c_2 = 2.39$
- $Z_2 = 3.60$
- Stop and reject the null hypothesis

Sequential monitoring of DBP: Summary

- The trial stops at the second interim analysis
- We reject the null hypothesis and conclude that subjects in the control group have a significantly higher DBP than subjects in the experimental condition
- We use a sample size of 160 subjects per group, which is a savings of 240 subjects per group compared to the fixed-sample design

Error Spending for One-sided Test or Two-sided Tests with an Inner Wedge

- To this point, we have only consider error spending in the context of two-sided tests with no inner-wedge
- Error spending functions can be easily extended to the case of a one-sided test or two-sided test with an inner wedge
- In this case, an error spending function is specified for both α and β

Error Spending for a One-sided Test

- For simplicity, we will only consider one-sided tests as two-sided tests with an inner wedge are very similar
- We now must specify an error spending function for both the type I and type II error
  - $f(t)$
  - $g(t)$

Error Spending for a One-sided Test

- f and g must both meet the requirement for an error spending function
  - $f(0) = g(0) = 0$
  - f and g are non-decreasing
  - $f(t) = \alpha$ and $g(1) = \beta$

Partitioning the Type I error

Type I error is partitioned as before

- $\pi_{1,1} = f\left(I_1/I_{max}\right)$
- $\pi_{1,k} = f\left(I_k/I_{max}\right) - f\left(I_{k-1}/I_{max}\right)$ for $k = 2, 3, \ldots$

Partitioning the Type II error

Type II error is partitioned in an identical fashion

- $\pi_{2,1} = g\left(I_1/I_{max}\right)$
- $\pi_{2,k} = g\left(I_k/I_{max}\right) - g\left(I_{k-1}/I_{max}\right)$ for $k = 2, 3, \ldots$

Critical Values

Critical Values are now specified as follows:

- The first critical values, a_1 and b_1 are defined as a_1 and b_1, such that:

$$P\left(Z_1 > b_1 | \delta = 0, I_1\right) = \pi_{1,1}$$

and

$$P\left(Z_1 < a_1 | \delta = \delta_a, I_1\right) = \pi_{2,1}$$

where δ_a is the pre-specified alternative hypothesis

Critical Values

- Subsequent critical values are defined as, a_k and b_k are defined as a_k and b_k, such that:

  $$P\left(a_1 < Z_1 < b_1, \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k > b_k | \delta = 0, I_1, \ldots, I_k\right) = \pi_{1,k}$$

  and

  $$P\left(a_1 < Z_1 < b_1, \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k < a_k | \delta = \delta_a, I_1, \ldots, I_k\right) = \pi_{2,k}$$

  where δ_a is the pre-specified alternative hypothesis

Critical Values

Intuitively, critical values are defined such that the correct amount of error is spent at each interim analysis conditional on continuing after the first $k-1$ interim analysis

One-sided Error Spending Functions and Power

- One-sided group sequential tests are designed to detect a specific power to detect a pre-determined alternative
- You must make assumptions about the number and timing of interim analyses in order to determine sample size/power for sequential tests using error spending functions
- These will likely change when conducting the trial
- In this case, it is usually not possible to determine $a_K = b_K$ that achieve the desired type-I and type-II error
- It is best to find b_K that achieves the desired type-I error rate and set $a_K = b_K$ realizing that the power will be a little off

Critical Values: Example

- Consider a two-sided test with the following properties
  - $\alpha = 0.05$
  - 90% power (i.e. $\beta = 0.10$) to detect an effect size of 0.20
  - This would require a fixed-sample size of 215
  - $f(t) = \alpha t^3$
  - $g(t) = \alpha t^3$
  - Assuming 5 equally spaced stopping times, we would require a maximum sample size of 226

Critical Value Example: First Interim Analysis

- The first interim analysis occurs after 23 subjects have been enrolled
- Upper boundary
  - $f\left(23/226\right) = f\left(0.102\right) = \alpha 0.102^3 = 0.00005$
  - $\pi_{1,1} = 0.00005$
  - $b_1 = \Phi\left(1 - \pi_{1,1}\right)^{-1} = 3.88$
- Lower boundary
  - $g\left(23/226\right) = f\left(0.102\right) = \beta 0.102^3 = 0.0001$
  - $\pi_{2,1} = 0.0001$
  - $a_1 = -2.75$
  - Note: $a_k \neq \Phi\left(1 - \pi_{1,1}\right)^{-1}$

Critical Value Example: Second Interim Analysis

- The second interim analysis occurs after 80 subjects have been enrolled
- Upper boundary
  - $f(80/226) = f(0.354) = \alpha 0.354^3 = 0.0022$
  - $\pi_{1,2} = 0.0022 - 0.00005 = 0.00215$
  - $b_2 = 2.85$
- Lower boundary
  - $g(80/226) = f(0.354) = \beta 0.354^3 = 0.0044$
  - $\pi_{2,2} = 0.0044 - 0.0001 = 0.0043$
  - $a_2 = -0.84$

Critical Value Example: Third Interim Analysis

- The third interim analysis occurs after 136 subjects have been enrolled
- Upper boundary
  - $f(136/226) = f(0.602) = \alpha 0.602^3 = 0.0109$
  - $\pi_{1,3} = 0.0109 - 0.0022 = 0.0087$
  - $b_3 = 2.33$
- Lower boundary
  - $g(136/226) = f(0.602) = \beta 0.602^3 = 0.0218$
  - $\pi_{2,3} = 0.0218 - 0.0044 = 0.0174$
  - $a_3 = 0.27$

Critical Value Example: Fourth Interim Analysis

- The fourth interim analysis occurs after 204 subjects have been enrolled
- Upper boundary
  - $f\,(204/226) = f\,(0.903) = \alpha 0.903^3 = 0.0368$
  - $\pi_{1,4} = 0.0368 - 0.0109 = 0.0259$
  - $b_4 = 1.83$
- Lower boundary
  - $g\,(204/226) = f\,(0.903) = \beta 0.903^3 = 0.0736$
  - $\pi_{2,4} = 0.0736 - 0.0218 = 0.0518$
  - $a_4 = 1.36$

Critical Value Example: Final Analysis

- The final analysis includes 226 subjects
- Upper boundary
  - $f(1) = 0.05$
  - $\pi_{1,5} = 0.05 - 0.0368 = 0.0132$
  - $b_5 = 1.69$
- Lower boundary
  - $a_5 = 1.69$
  - Final power is 90.01/

One-sided sequential monitoring of SBP using an error spending function

Consider one-sided sequential monitoring of SBP at 24 months in the Mr Fit study using the error spending approach

- $\alpha = 0.05$
- Assume that σ is known and equal to 14
- We would like 90% power to detect a significant difference of 5 mmHg
- This will require a fixed-sample size of 135 subjects per group

One-sided sequential monitoring of SBP using an error spending function

- We will use the Kim and DeMets error spending function with $\rho = 3$ for both α and β spending

$$f(t) = \alpha t^3$$

and

$$g(t) = \beta t^3$$

- A sample size of 142 subjects per group is required to achieve 90% assuming five equally spaced interim analyses

Sequential monitoring of SBP: first interim analysis

The first interim analysis uses the first 15 subjects per group

- Upper boundary
  - $f(15/142) = f(0.106) = 0.00005$
  - $\pi_{1,1} = 0.00005$
  - $b_1 = 3.85$
- Lower boundary
  - $g(15/142) = g(0.106) = 0.0001$
  - $\pi_{2,1} = 0.0001$
  - $a_1 = -2.70$
- $Z_1 = .365$
- Continue the trial

Sequential monitoring of SBP: second interim analysis

The second interim analysis uses the first 45 subjects per group

- Upper boundary
  - $f(45/142) = f(0.317) = 0.0016$
  - $\pi_{1,2}$ = 0.0016 - 0.00005 = 0.00155
  - $b_2 = 2.95$
- Lower boundary
  - $g(45/142) = g(0.317) = 0.0032$
  - $\pi_{2,2}$ = 0.0032 - 0.0001 = 0.0031
  - $a_2 = -1.04$
- $Z_2 = 1.71$
- Continue the trial

Sequential monitoring of SBP: third interim analysis

The third interim analysis uses the first 45 subjects per group

- Upper boundary
  - $f(70/142) = f(0.493) = 0.0060$
  - $\pi_{1,3}$ = 0.0060 - 0.0016 = 0.0044
  - $b_3 = 2.56$
- Lower boundary
  - $g(70/142) = g(0.493) = 0.0120$
  - $\pi_{2,3}$ = 0.0120 - 0.0032 = 0.0088
  - $a_3 = -0.19$
- $Z_3 = 0.73$
- Continue the trial

Sequential monitoring of SBP: fourth interim analysis

The fourth interim analysis uses the first 115 subjects per group

- Upper boundary
  - $f(115/142) = f(0.810) = 0.0266$
  - $\pi_{1,4}$ = 0.0266 - 0.0060 = 0.0206
  - $b_4 = 1.97$
- Lower boundary
  - $g(115/142) = g(0.810) = 0.0531$
  - $\pi_{2,4}$ = 0.0531 - 0.0120 = 0.0411
  - $a_4 = 1.06$
- $Z_4 = 2.38$
- Stop and reject the null hypothesis

Sequential monitoring of SBP: Summary

- The trial stops at the fourth interim analysis
- We reject the null hypothesis and conclude that subjects in the control group have a significantly higher SBP than subjects in the experimental condition
- We use a sample size of 115 subjects per group, which is a savings of 10 subjects per group compared to the fixed-sample design
- Had the trial reached full enrollment, we would have had slightly more than 90% power (90.13%)

One-sided sequential monitoring of DBP using an error spending function

Consider one-sided sequential monitoring of DBP at 24 months in the Mr Fit study using the error spending approach

- $\alpha = 0.05$
- Assume that σ is known and equal to 8
- We would like 90% power to detect a significant difference of 2 mmHg
- This will require a fixed-sample size of 275 subjects per group

One-sided sequential monitoring of DBP using an error spending function

- We will use the Kim and DeMets error spending function with $\rho = 3$ for both α and $\rho = 2$ for β spending

$$f(t) = \alpha t^3$$

and

$$g(t) = \beta t^2$$

- A sample size of 296 subjects per group is required to achieve 90% assuming five equally spaced interim analyses

Sequential monitoring of SBP: first interim analysis

The first interim analysis uses the first 15 subjects per group

- Upper boundary
  - $f(150/296) = f(0.507) = 0.0065$
  - $\pi_{1,1}$ = 0.0065
  - $b_1 = 2.48$
- Lower boundary
  - $g(150/296) = g(0.507) = 0.0257$
  - $\pi_{2,1}$ = 0.0257
  - $a_1 = 0.21$
- $Z_1 = 3.25$
- Stop and reject the null hypothesis

Sequential monitoring of DBP: Summary

- The trial stops at the first interim analysis
- We reject the null hypothesis and conclude that subjects in the control group have a significantly higher DBP than subjects in the experimental condition
- We use a sample size of 150 subjects per group, which is a savings of 146 subjects per group compared to the fixed-sample design
- Had the trial reached full enrollment, we would have had slightly more than 90% power (91%)