Chapter 126

Group-Sequential Tests for One Proportion in a Fleming Design

Introduction

This procedure computes power and sample size for the single-arm group-sequential (multiple-stage) designs of Fleming (1982). These designs are used for testing one-sided hypotheses about the proportion responding. They are often used in Phase II clinical trials determine whether a drug or regimen has sufficient activity against disease to warrant more extensive study and development. In a group-sequential design, the patients are divided into two or more stages. At the end of each stage, an interim statistical test is made to determine if the next stage should be conducted.

Technical Details

Phase I clinical trials are designed to provide information about the maximum tolerated dose levels of a treatment. They consist of three to six patients at each dose level and provide little information about the effectiveness of the treatment.

Phase II trials obtain initial estimates of the degree of treatment activity. A patient's response may be measured by the decrease in the size of a tumor. For example, a patient may be considered to have responded to treatment if the tumor shrinks by 50% or more. There is no control group in these designs. Rather, the purpose of the trial is to determine if the drug shows enough activity against disease to warrant a full-scale, phase III clinical trial.

Fleming Multiple Test Procedure

Fleming (1982) gave a formal method for determining acceptance and rejection points in a single-arm group-sequential design. Fleming's design is for testing $H_0: p \le p_0$ versus $H_A: p > p_1$. The power of this test is computed using the formulas in Schultz, Nichol, Elfring, and Weed (1973).

In his paper, Fleming gave results for two and three stage designs. His designs allow early stopping of the study if interim tests results are extreme, either for or against the treatment. The final test uses the standard error-rate probabilities.

Group-Sequential Tests for One Proportion in a Fleming Design

In the discussion below, p is the true proportion responding to the treatment in the population and p_0 is the largest response proportion which, if true, clearly implies that the treatment does not warrant further study. The parameter p_0 is sometimes called the response rate of a poor treatment. For a new anti-tumor drug, this may be set to a small value such as 0.05 or 0.10.

The power of this design is computed at $p = p_1$, where p_1 is the smallest response proportion which, if true, clearly implies that the treatment does warrant further study. The parameter p_1 is sometimes called the response rate of a good treatment. For a new anti-tumor drug, this may be set to a value such as 0.25 or 0.30.

This group-sequential design with K stages is defined as follows

n_g is the number of individuals recruited in stage g. $N_g = n_1 + n_2 + \cdots + n_g$

s_g is the number of positive responses in stage g. $S_g = s_1 + s_2 + \cdots + s_g$

a_g is the acceptance point for stage g.   If $S_g \leq a_g$, stop and reject H_A

r_g is the rejection point for stage g.   If $S_g \geq r_g$, stop and reject H_0

                                               If $a_g < S_g < r_g$ continue to stage $g + 1$

Each interim test, based on the test statistic Y, is defined as follows.

$$Y_g(p) = \frac{S_g - \sum_{i=1}^{g} n_i p}{\sqrt{N_g p(1-p)}}$$

At interim test g, reject H_0 when

$$Y_g(p_0)\sqrt{\frac{N_g}{N}} > z_{1-\alpha}$$

Similarly for $g < K$, reject H_A when

$$Y_g(p_1)\sqrt{\frac{N_g}{N}} < -z_{1-\alpha}$$

or at test K if rejection of H_0 does not occur.

The acceptance and rejection values for each stage, a_g and r_g, are found using

$$r_g = \left[\sum_{i=1}^{g} n_i p_0 + z_{1-\alpha}\sqrt{N_K p_0(1-p_0)}\right]^* + 1, \qquad g = 1, \ldots, K$$

$$a_g = \begin{cases} \left[\sum_{i=1}^{g} n_i \tilde{p}_A + z_{1-\alpha}\sqrt{N_K \tilde{p}_A(1-\tilde{p}_A)}\right]^* & \text{if } g = 1, \ldots, K-1 \\ r_K - 1 & \text{if } g = K \end{cases}$$

$$\tilde{p}_A = \frac{\left(\sqrt{N_K p_0} + z_{1-\alpha}\sqrt{1-p_0}\right)^2}{N_K + z_{1-\alpha}^2}$$

where $[x]^*$ means to round to the nearest integer (note that ties which are exactly between two integers, such as 3.5, are rounded to the closest integer furthest from zero).

Schultz's Calculations

Schultz, Nichol, Elfring, and Weed (1973) give formulas for computing the type-I and type-II error probabilities of α and β, respectively.

Using the above notation, the probability of accepting H_0 at stage g is

$$L_g(p) = \sum_{m=a_{g-1}+1}^{a_g} C_g(m, p)$$

where

$$C_g(m, p) = \sum_{j=\max(a_{g-1}+1, m-n_g)}^{\min(r_{g-1}-1, m)} C_{g-1}(j, p) \, B_g(m-j, p)$$

$$B_g(x, p) = \binom{n_g}{x} p^x (1-p)^{n_g - x}$$

Likewise, the probability of reject H_0 at stage g is

$$R_g(p) = \sum_{m=r_g}^{r_{g-1}+n_g-1} C_g(m, p)$$

Using the above, the exact values of alpha and power are

$$\alpha = 1 - \sum_{g=1}^{K} L_g(p_0)$$

$$Power = 1 - \beta = \sum_{g=1}^{K} R_g(p_1)$$

The average sample number (ASN) may be computed for both p_0 and p_1 using

$$ASN(p) = \sum_{g=1}^{K} N_g \{L_g(p) + R_g(p)\}$$

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

Select whether you want to solve for sample size or power.

- **Sample Size**

  A search is made of a range of sample sizes looking for the smallest sample size that meets the type-I and type-II error requirements. This design will usually have the smallest expected sample size as well.

  Only the optimum design is reported.

- **Power**

  The power is calculated for each design specified. This option is often run to see the results for a range of sample sizes.

Power and Alpha

Target Power (Solve For = Sample Size Only)

This is the power desired. It will seldom be achieved exactly because the test statistic, the number of successes, only takes on integer values.

Power is the probability of rejecting the null hypothesis when it is false. It is calculated at $P = P1$ using the binomial probability distribution. Power is equal to $1 - \beta$, so specifying power implicitly specifies β. Beta is the probability of obtaining a false negative with the statistical test. That is, it is the probability of accepting a false null hypothesis.

The valid range is 0 to 1. Usually, a target power between 0.8 and 0.95 is selected.

You can enter a single value such as *0.90* or a series of values such as *0.70 0.80 0.90.*

When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

Target Alpha

This is the alpha level desired. It will seldom be achieved exactly because the test statistic, the number of successes, only takes on integer values.

Alpha is the probability of obtaining a false positive with the statistical test. That is, it is the probability of rejecting a true null hypothesis. It is calculated from the binomial distribution using $P = P0$. The null hypothesis is that $P \leq P0$.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.10. Alpha is commonly set to 0.05. However, since this is a one-sided test, some set it to 0.025.

You can enter a single value such as *0.05* or a series of values such as *0.025 0.05 0.10* or *0.025 to 0.10 by 0.025.*

When a series of values is entered, **PASS** will generate a separate calculation result for each value of the series.

Effect Size – Proportion Responding

P0 (Maximum Poor P)

P0 is the largest value of P that still results in the conclusion that the treatment is poor and need not be studied further. If the response proportion is this value or smaller, the treatment is said to be futile and will not be pursued further.

The strict range is $0 < P0 < P1 < 1$. However, usually you would use values less than 0.10. Often, 0.05 is used.

You can enter a single value such as *0.05* or a series of values such as *0.05 0.10 0.15* or *0.05 to 0.2 by 0.05*.

P1 (Minimum Good P)

This is the response proportion of a good drug. If the true proportion responding to the treatment is greater than or equal to this amount, study of the treatment would be recommended.

P1 is the smallest value of P that still results in the conclusion that the treatment is good and should be studied further. If the response proportion is this value or larger, the treatment is said to be effective.

Note that P1 is the value of P at which the power is calculated. However, it does not define the alternative hypothesis. The alternative hypothesis is the $P > P0$. The range is $0 < P0 < P1 < 1$.

You can enter a single value such as 0.05 or a series of values such as 0.05 0.10 0.15 or 0.05 to 0.2 by 0.05.

Stages

Number of Stages (Looks)

This is the number of stages (often called looks) at which statistical tests are conducted. This number includes the final test, so it is the number of intermediate tests plus one. Fleming (1982) stated that for Phase II clinical trials, only two or three stages are usually used. However, the possible choices are from 1 to 20.

Sample Size (When Solve For = *Sample Size*)

Sample Size Search Method

This option indicates how a sample size meeting the alpha and power criterion are searched for.

Possible choices are

- **Use the one-stage sample size**

  Calculate and use the sample size of a single-stage study having the same parameter settings. Fleming (1982) proposed this as a reasonable choice.

- **Find the smallest between one-stage N ± m**

  Calculate the single-stage sample size: *S*. Calculate the actual alpha and power from *S - m* to *S + m*. Stop the search and report the first sample size that meets the alpha and power requirements.

- **Find smallest between N1 and N2**

  Calculate the actual alpha and power from *N1* to *N2*. Stop the search and report the first sample size that meets the alpha and power requirements.

m

This is the number of units below the single-stage sample size (*S*) that a search is conducted. The search goes from *S - m* to *S + m*. The search is stopped when a sample size is found that meets the alpha and power requirements.

For example, suppose the single-stage sample size is found to be 35 and the value of *m* is 10. The search for a valid sample size would be from 25 to 45. The search would be terminated as soon as a sample size that meets the alpha and power requirements is found.

N1

This is the starting value for a search to find the smallest sample size that meets the alpha and power requirements. The search is conducted from N1 and N2.

N1 should be at least twice the number of stages.

N2

This is the ending value for a search to find the smallest sample size that meets the alpha and power requirements. The search is conducted from N1 and N2.

N2 should be greater than or equal to N1. It can be a large number since it is an upper bound that is usually not reached.

Stage Sample Percentages

This option indicates how the stage sample size percentages are entered. Possible choices are

- **Use equal stage sample size percentages**

  The total sample size N is divided equally among the stages. Fleming (1982) recommended equal allocation.

- **Enter a percentage for each stage**

  Enter an individual stage percentage for each stage below.

Percent of N (for each Stage)

This is the percentage of the total sample size that is allocated to this stage. If these values do not sum to exactly 100, they are rescaled so that they do.

The individual stage samples sizes are determined by taking this percentage of the N. If these values are represented by *p1*, *p2*, *p3*, ..., the stage sample sizes are calculated as follows:

$$n1 = [N \times p1 / 100]$$

$$n2 = [N \times p2 / 100]$$

$$n3 = [N \times p3 / 100]$$

where the operator, [X] means the closest integer to X, e.g. [3.1]=3 and [5.6] = 6.

For example, suppose there are three stages and these percentages are set to 25, 25, and 50. If N is 36, the resulting stage sample sizes are 9, 9, and 18.

Sample Size (When Solve For = *Power*)

Stage Sample Allocation

This option indicates how the sample size is entered. Possible choices are

- **Enter N (total sample size) and use equal stage sample sizes**

  Enter the total sample size N and divide it equally among the stages. Fleming (1982) recommended equal allocation.

- **Enter N (total sample size) and a percentage for each stage**

  Enter the total sample size N and also enter the individual stage percentages below.

- **Enter base stage sample size and use equal stage multipliers**

  Enter a base sample size to be used for all stages. The total sample size will be this value times the number of stages. Fleming (1982) recommended equal allocation.

- **Enter base stage sample size and a multiplier for each stage**

  Enter a base sample size. Find each stage sample size be multiplying this base value times a stage multiplier entered below.

- **Enter a sample size for each stage**

  Enter the stage sample sizes directly.

N (Total Sample Size)

This is the total sample size for all stages. One or more values, separated by blanks or commas, may be entered. A separate analysis is performed for each value listed here.

If equal stage sizes have been specified, this value is divided equally among the stages.

If unequal stage sizes have been specified, the individual stage samples sizes are determined by multiplying this value times the corresponding Percent of N value entered below.

If the percent values are represented by $p1$, $p2$, $p3$, ..., and this value is represented by N, the stage sample sizes are calculated as follows:

$$n1 = [N \times p1 \, / \, 100]$$
$$n2 = [N \times p2 \, / \, 100]$$
$$n3 = [N \times p3 \, / \, 100]$$

where the operator, [X] means the closest integer to X, e.g. [3.1]=3 and [5.6] = 6.

Percent of N for each Stage

This is the ending value for a search to find the smallest sample size that meets the alpha and power requirements. The search is conducted from N1 and N2.

N2 should be greater than or equal to N1. It can be a large number since it is an upper bound that is usually not reached.

Group-Sequential Tests for One Proportion in a Fleming Design

Base Stage Sample Size

This is the base sample size per stage. One or more values, separated by blanks or commas, may be entered. A separate analysis is performed for each value listed here.

If equal stage sizes are specified, this is the sample size for all stages.

If separate multipliers are specified, the individual stage samples sizes are determined by multiplying this value times the corresponding Multiplier value entered below.

If the Multiplier numbers are represented by *m1*, *m2*, *m3*, ... and this value is represented by *n*, the stage sample sizes *n1, n2, n3*, ... are calculated as follows:

$$n1 = [Base \times m1]$$

$$n2 = [Base \times m2]$$

$$n3 = [Base \times m3]$$

where the operator, [X] means the closest integer to X, e.g. [3.1]=3 and [3.6] = 4.

Multiplier for each Stage

The sample size multiplier for this stage is entered here. The sample size for this stage is this value times the Base Stage Sample Size.

The actual range for this value is any value greater than zero. Typical values of this parameter are 0.5, 1, and 2.

If the Multiplier numbers are represented by *m1*, *m2*, *m3*, ... and this value is represented by *n*, the stage sample sizes *n1, n2, n3*, ... are calculated as follows:

$$n1 = [Base \times m1]$$

$$n2 = [Base \times m2]$$

$$n3 = [Base \times m3]$$

where the operator, [X] means the closest integer to X, e.g. [3.1]=3 and [3.6] = 4.

Sample Size for Stage

The individual stage samples are entered here, one per box. Strictly speaking, the range is 1 to N. However, it is hard to imagine a case where these values are less than five.

Example 1 – Calculating the Power

Suppose you want to look at two-stage Fleming designs between *N* = 50 and 70, *Alpha* = 0.05, *P0* = 0.05, and *P1* = 0.15. This would be set up as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Tests for One Proportion in a Fleming Design** procedure. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Target Alpha	**0.05**
P0 (Maximum Poor P)	**0.05**
P1 (Minimum Good P)	**0.15**
Number of Stages..................................	**2**
Stage Sample Allocation	**Enter N (total sample size and use equal stage sample sizes)**
N (Total Sample Size)............................	**50 to 70 by 2**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Summary Results of Fleming's Group-Sequential Design

Scenario	Actual Power	N	Ave N at P=P0 ASN(P0)	Ave N at P=P1 ASN(P1)	Maximum Poor P0	Minimum Good P1	Actual Alpha	Target Alpha
1	0.7806	50	42.9	41.6	0.050	0.150	0.0391	0.0500
2	0.8116	52	44.9	42.5	0.050	0.150	0.0460	0.0500
3	0.8391	54	47.0	43.4	0.050	0.150	0.0536	0.0500
4	0.8632	56	49.0	44.1	0.050	0.150	0.0619	0.0500
5	0.7964	58	51.1	44.8	0.050	0.150	0.0318	0.0500
6	0.8240	60	53.1	45.5	0.050	0.150	0.0370	0.0500
7	0.8373	62	44.8	45.0	0.050	0.150	0.0414	0.0500
8	0.8595	64	46.7	45.7	0.050	0.150	0.0475	0.0500
9	0.8730	66	49.2	52.6	0.050	0.150	0.0453	0.0500
10	0.8911	68	51.2	53.4	0.050	0.150	0.0518	0.0500
11	0.9068	70	53.2	54.1	0.050	0.150	0.0588	0.0500

References

Fleming, Thomas R. 1982. 'One-Sample Multiple Testing Procedure for Phase II Clinical Trials.' Biometrics, 38, pages 143-151.

Schultz, J.R., Nichol, F.R., Elfring, G.L., and Weed, S.D. 1973. 'Multiple stage procedures for drug screening.' Biometrics, 29, pages 293-300.

Jennison, C., and Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC. New York.

Report Definitions
H0 is the null hypothesis that P ≤ P0.
P is the probability that an individual responds favorably.
P0 is the largest response probability of a POOR treatment.
P1 is the smallest response probability of a GOOD treatment.
Target Power is the desired probability of rejecting a false H0.
Actual Power is the achieved probability of rejecting a false H0.
Target Alpha is the desired probability of rejecting a true H0.
Actual Alpha is the Alpha level actually achieved by this design.
N is the total sample size, if all stages are completed.
ASN(P0) is the average total sample size, if P = P0.
ASN(P1) is the average total sample size, if P = P1.

Summary Statements
A Fleming 2-stage design with 50 subjects tests the null hypothesis that the response rate is
less than or equal to 0.050 versus the alternative that it is greater. At each stage, the
cumulative number of successes, R, is compared to the acceptance (Ag) and rejection (Rg)
numbers for that stage. If R ≤ Ag, the experiment is stopped for futility. If R ≥ Rg, it is
stopped for efficacy. The type I error rate is 0.0391. The power, calculated at a response rate
of 0.150, is 0.7806. The acceptance numbers are [0, 5]. The rejection numbers are [5, 6]. The
sample sizes added at each stage are [25, 25].

This report shows the actual power and alpha for all the designs considered, one row at a time. Note that the smallest sample size that meets the error-rate criterion is 52. It is interesting to look down the values of N and watch the power values. We note that at N = 58, the power criterion is not met—even though it was at 52. Also, we note that N = 52, 54, 68, and 70 do not meet the target alpha requirements. This emphasizes why one must consider a range of values around a selected one because seldom is the final sample size equal to the planned on.

Stage Results of Fleming's Group-Sequential Design for Scenario 1
N = 50 P0 = 0.050 P1 = 0.150 Alpha = 0.0391 Power = 0.7806

Stage g	Stage Sample Size Ng	Cumulative Sample Size Σ(Ng)	Stage Sample Size Percent	Cumulative Sample Size Σ(Percent)	Accept H0 if R ≤ Ag Ag	Reject H0 if R ≥ Rg Rg
1	25	25	50.0	50.0	0	5
2	25	50	50.0	100.0	5	6

Report Definitions
Scenario is a single combination of parameter settings.
The value g is the stage number.
Stage Sample Size, Ng, is the sample size in stage g.
Cumulative Stage Sample Size is the total sample size through stage g.
Stage Sample Size Percent is the percent of the total sample size occuring in stage g.
Cumulative Sample Size Percent the percentage sample size through stage g.
R is the cumulative number of successes through stage g.
If Ag or fewer successes occur through stage g, the study is stopped for futility.
If Rg or more successes occur through stage g, the study is stopped for efficacy (H0 is rejected).

This report shows the individual stage statistics. Of special interest are the acceptance and rejection numbers, *Ag* and *Rg*. For stage 1, the acceptance and rejection numbers are 0 and 5. For stage 2, they are 5 and 6.

Example 2 – Calculating the Sample Size

Suppose you want to find a two-stage Fleming design for *Alpha* = 0.05, *P0* = 0.05, and *P1* = 0.15. This is set up as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Tests for One Proportion in a Fleming Design** procedure. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Target Power	**0.80**
Target Alpha	**0.05**
P0 (Maximum Poor P)	**0.05**
P1 (Minimum Good P)	**0.15**
Number of Stages....................................	**2**
Sample Size Search Method	**Find the smallest between the one-stage N ± m**
m...	**10**
Stage Sample Percentages	**Use equal stage sample size percentages**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Summary Results of Fleming's Group-Sequential Design

Scenario	Actual Power	Target Power	N	Ave N at P=P0 ASN(P0)	Ave N at P=P1 ASN(P1)	Maximum Poor P0	Minimum Good P1	Actual Alpha	Target Alpha
1	0.8116	0.8000	52	44.9	42.5	0.050	0.150	0.0460	0.0500

Stage Results of Fleming's Group-Sequential Design for Scenario 1
N = 52 P0 = 0.050 P1 = 0.150 Alpha = 0.0460 Power = 0.8116

Stage g	Stage Sample Size Ng	Cumulative Sample Size Σ(Ng)	Stage Sample Size Percent	Cumulative Sample Size Σ(Percent)	Accept H0 if R ≤ Ag Ag	Reject H0 if R ≥ Rg Rg
1	26	26	50.0	50.0	0	5
2	26	52	50.0	100.0	5	6

PASS has calculated the optimum sample size to be N = 52 which matches what we found in Example 1.

Example 3 – Validation using Fleming (1982)

Fleming (1982) page 147 in his Table 2 presents designs for several scenarios. The third row of the table gives the results for a three-stage design in which the individual stage sample sizes are 10, 5, and 5. The value of $P0$ is 0.05, $P1 = 0.20$, and *alpha* = 0.05. The values of Ag are 0, 1, and 3. The values of Rg are 3, 3, and 4. The actual power is 0.651 and the actual alpha is 0.038.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Tests for One Proportion in a Fleming Design** procedure. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Target Alpha	**0.05**
P0 (Maximum Poor P)	**0.05**
P1 (Minimum Good P)	**0.20**
Number of Stages...................................	**3**
Stage Sample Allocation	**Enter a sample size for each stage**
Stage 1 Sample Size	**10**
Stage 2 Sample Size	**5**
Stage 3 Sample Size	**5**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Summary Results of Fleming's Group-Sequential Design

Scenario	Actual Power	Ave N at P=P0 NASN(P0)	Ave N at P=P1 ASN(P1)	Maximum Poor P0	Minimum Good P1	Actual Alpha	Target Alpha
1	0.6506	20 12.6	13.9	0.050	0.200	0.0383	0.0500

Stage Results of Fleming's Group-Sequential Design for Scenario 1
N = 20 P0 = 0.050 P1 = 0.200 Alpha = 0.0383 Power = 0.6506

Stage g	Stage Sample Size Ng	Cumulative Sample Size Σ(Ng)	Stage Sample Size Percent	Cumulative Sample Size Σ(Percent)	Accept H0 if R ≤ Ag Ag	Reject H0 if R ≥ Rg Rg
1	10	10	50.0	50.0	0	3
2	5	15	25.0	75.0	1	3
3	5	20	25.0	100.0	3	4

You can see that all values match.