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 to 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 $H_0: p \leq p_0$ 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.

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.

© NCSS, LLC. All Rights Reserved.

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 $S_g \le a_g$, stop and reject H_A
r_g is the rejection point for stage *g*.	If $S_g \ge r_g$, 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).

© NCSS, LLC. All Rights Reserved.

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

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

© NCSS, LLC. All Rights Reserved.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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**

Output

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

Summary Results of Fleming's Group-Sequential Design

Solve For: Power

Scenario	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	0.05	0.15	0.0391	0.05
2	0.8116	52	44.9	42.5	0.05	0.15	0.0460	0.05
3	0.8391	54	47.0	43.4	0.05	0.15	0.0536	0.05
4	0.8632	56	49.0	44.1	0.05	0.15	0.0619	0.05
5	0.7964	58	51.1	44.8	0.05	0.15	0.0318	0.05
6	0.8240	60	53.1	45.5	0.05	0.15	0.0370	0.05
7	0.8373	62	44.8	45.0	0.05	0.15	0.0414	0.05
8	0.8595	64	46.7	45.7	0.05	0.15	0.0475	0.05
9	0.8730	66	49.2	52.6	0.05	0.15	0.0453	0.05
10	0.8911	68	51.2	53.4	0.05	0.15	0.0518	0.05
11	0.9068	70	53.2	54.1	0.05	0.15	0.0588	0.05

P	The probability that an individual responds favorably.
H0	The null hypothesis that P ≤ P0.
Power	The probability of rejecting a false H0.
N	The total sample size if all stages are completed.
ASN(P0)	The average total sample size if P = P0.
ASN(P1)	The average total sample size if P = P1.
P0	The largest response probability of a POOR treatment.
P1	The smallest response probability of a GOOD treatment.

© NCSS, LLC. All Rights Reserved.

Actual Alpha	The Alpha level actually achieved by this design.
Target Alpha	The desired probability of rejecting a true H0.

Summary Statements

A Fleming 2-stage single-arm trial design will be used to test whether the proportion responding (P) warrants continuation to the next phase (H0: P ≤ 0.05 versus H1: P ≥ 0.15). A Type I error rate (α) of 0.05 will be used. At each stage, the cumulative number of successes is compared to the null hypothesis acceptance cutoff value and rejection cutoff value. If the cumulative number of successes is less than or equal to the acceptance cutoff value, the trial is stopped for futility. If the cumulative number of successes is greater than or equal to the rejection cutoff value, the trial is stopped for efficacy. For a good treatment minimum response rate of 0.15 with a final stage total sample size of 50 subjects, the power is 0.7806. The successive acceptance cutoff values are [0, 5]. The successive rejection cutoff values are [5, 6]. The numbers of subjects added at each stage are [25, 25].

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.

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 go down through 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 should consider a range of values around a target value.

Stage Results of Fleming's Group-Sequential Design for Scenario 1

N = 50 P0 = 0.05 P1 = 0.15 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	50	0	5
2	25	50	50	100	5	6

Scenario	A single combination of parameter settings.
g	The stage number.
Ng	The sample size in stage g.
Σ(Ng)	The cumulative sample size through stage g.
Stage Sample Size Percent	The percent of the total sample size occurring in stage g.
Σ(Percent)	The cumulative sample size percentage through stage g.
R	The cumulative number of successes through stage g.
Ag	If Ag or fewer successes occur through stage g, the study is stopped for futility.
Rg	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.

© NCSS, LLC. All Rights Reserved.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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.

Summary Results of Fleming's Group-Sequential Design

Solve For: Sample Size

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	0.8	52	44.9	42.5	0.05	0.15	0.046	0.05

Stage Results of Fleming's Group-Sequential Design for Scenario 1

N = 52 P0 = 0.05 P1 = 0.15 Alpha = 0.046 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	50	0	5
2	26	52	50	100	5	6

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

© NCSS, LLC. All Rights Reserved.

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 *P*0 is 0.05, *P*1 = 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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.

Summary Results of Fleming's Group-Sequential Design

Solve For: Power

Scenario	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	0.6506	20	12.6	13.9	0.05	0.2	0.0383	0.05

© NCSS, LLC. All Rights Reserved.

Stage Results of Fleming's Group-Sequential Design for Scenario 1

N = 20 P0 = 0.05 P1 = 0.2 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	50	0	3
2	5	15	25	75	1	3
3	5	20	25	100	3	4

You can see that all values match.

© NCSS, LLC. All Rights Reserved.