Chapter 125

Two-Stage Designs for Tests of One Proportion (Simon)

Introduction

This module finds two-stage designs for exact tests of a single proportion that meet the error rate (type-I and type-II) criterion and minimize the expected sample size. An algorithm, presented by Simon (1989), finds the designs with the minimum N (*minimax*) and the minimum expected N (*optimum*). Extending Simon's work, Jung, Lee, Kim, George (2004) discusses other designs which are optimum from a Bayesian point of view which they call *admissible* designs.

In a two-stage design, the subjects are divided into two groups or stages. At the completion of the first stage, an interim analysis is made to determine if the second stage should be conducted. If the number of patients responding is greater than a certain amount, the second stage is conducted. Otherwise, it is not.

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 treatment shows enough activity against disease to warrant a full-scale, phase III clinical trial.

Let *P0* be the largest response proportion which, if true, clearly implies that the treatment does not warrant further study. *P0* is sometimes called the response rate of a *poor* treatment. For a new anti-tumor drug, this may be set to 0.10.

Let *P1* be the smallest response proportion which, if true, clearly implies that the treatment does warrant further study. *P1* is sometimes called the response rate of a good treatment. For a new anti-tumor drug, this may be set to 0.30.

A statistical test of hypothesis may be conducted to test the null hypothesis that $P \leq P0$ versus the alternative hypothesis that $P \geq P1$ (P is the true proportion responding to the treatment in the population). Let α be the probability of rejecting the null hypothesis when it is true. Let β be the probability of rejecting the alternative hypothesis when $P = P1$.

A phase II design can be represented by four numbers: *N1*, *R1*, *N*, and *R*. *N1* is the sample size in the first stage. *R1* is the critical value in the first stage. If *R1* or fewer responses occur in the *N1* patients, the drug is rejected. *N* is the combined sample size for both the first and second stages. *R* is the critical value in the combined sample. If *R* or fewer of the *N* patients respond, the drug is rejected.

The expected (or average) sample size of this design is

$$E(N) = N1 + (1 - PET)(N - N1)$$

where *PET* is the probability of early termination of the study computed at *P* = *P*0.

The probability of rejecting a drug with success probability *P* can be found using the binomial distribution. The formulation is

$$Pr(\text{reject}|P, N1, R1, R, N) = B(R1|P, N1) + \sum_{X=R1+1}^{\min(N1,R)} b(X|P, N1)B(R - X|P, N - N1)$$

where

$$b(X|P, N) = \frac{N!}{X!\,(N - X)!}P^X(1 - P)^{N-X}$$

$$B(X|P, N) = \sum_{r=0}^{X} b(R|P, N)$$

The two error rate constraints are

$$Pr(\text{reject}|P0, N1, R1, R, N) \geq 1 - \alpha$$

and

$$Pr(\text{reject}|P1, N1, R1, R, N) \leq \beta$$

Optimum Designs

Simon (1989) defines two optimum designs. The first is the *minimax* design which is the design with the smallest *N* that obeys both error rate constraints. The second is the *optimum* design. It has the minimum expected *N*. These designs are found through a search of all possible designs. This search may take several minutes to complete.

Occasionally, neither the minimax nor the optimum design is useful. For example, N1 may be too small or too large. Jung, Lee, Kim, George (2004) developed other designs which are optimum from a Bayesian point of view which they call *admissible* designs. These are a compromise between the two Simon designs.

The optimum design minimizes the average sample size, *E*(*N*), while meeting the error rate constraints.

Designs Other Than Optimal

The optimal design minimizes the average sample size. There are examples where a less-than optimal design may be more desirable. For example, suppose the optimal design were *N*1 = 5 and *N* = 25. This design is poor because only 5 patients are obtained during the first stage, but 20 are needed during the second stage. Most researchers would rather have more balance in the sample sizes of the two stages. Because of this, the actual optimal design may be rejected on other grounds. There are often designs that are near optimal but are much more balanced.

Design Flexibility

Dealing with sequential designs is complicated. It may be difficult to achieve exactly the number of patients proscribed for each phase. However, it should be remembered that the validity of the probability statements depends on the sample size requirements being met exactly. This is because the interpretation of an error rate probability statement is for repeated studies conducted in exactly the same way. We envision that if many studies of the same drug are conducted using the specific sampling plan *N1*, *R1*, *N*, *R* when *P* = *P0*, a proportion α of them will be falsely terminated due to chance occurrences.

The interpretation of the error rates is for a large number of identical studies in which the sampling plan is identical and as proscribed. If the sampling plan is allowed to vary, this interpretation is invalid. Of course, the degree of possible error in interpretation depends on the degree to which the sampling plan is changed. We recognize that when dealing with human subjects, flexibility must be maintained. However, the scientist must also recognize that when the sampling plan is changed, the exact probability statements can no longer be calculated.

Custom Search - Minimum N (Combined Sample Size)

N is the combined sample size of the two stages of the design. This parameter sets the minimum value of *N* that is used during the search. The optimum value of *N* must be between N Min and N Max, or it will not be found.

The keyword MIN indicates that the value used is the minimum of the smallest sample size from a single stage design and *MIN*2 where *MIN*2 is calculated using

$$MIN2 = \frac{p_0 + p_1}{2}\left(1 - \frac{p_0 + p_1}{2}\right)\left[\frac{z_{1-\alpha} + z_{1-\beta}}{p_1 - p_0}\right]^2$$

Since it is unlikely that the two-stage sample size will be less MIN, this provides a reasonable starting point for a search for *N*. However, experience has shown that you should use a small number such as 2 to ensure that you obtain the optimum.

Example 1 – Finding the Optimal Designs

Suppose a design is wanted for the case alpha = 0.05, power = 0.80, *P0* = 0.1, and *P1* = 0.25. 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

Search Type ..... **Automatic**
Power (Probability of Rejecting H0 P ≤ P0 | P = P1)..... **0.80**
Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0)..... **0.05**
P0 (Maximum Response Rate of a Poor Treatment)..... **0.10**
P1 (Minimum Response Rate of a Good Treatment)..... **0.25**
Multiplier for Upper Search Limit of N **1.25**

Output

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

Numeric Reports

Designs

Solve For: Sample Size
Beta Target: 0.2
N Search Limit Multiplier: 1.25

	Power		Sample Size				Response Rate			Alpha			Critical Value	
Design Type	Target	Actual	N1	N2	N	E(N)	Maximum of a Poor Treatment P0	Minimum of a Good Treatment P1	Probability of Early Termination PET	Target	Actual	Actual Beta	R1	R2
Single Stage	0.8	0.81805	40		40		0.1	0.25		0.05	0.04190	0.18195	7	
Min N	0.8	0.80319	22	18	40	28.84	0.1	0.25	0.62004	0.05	0.03980	0.19681	2	7
Admissible	0.8	0.80289	15	26	41	26.72	0.1	0.25	0.54904	0.05	0.04298	0.19711	1	7
Admissible	0.8	0.80416	14	28	42	25.63	0.1	0.25	0.58463	0.05	0.04641	0.19584	1	7
Min E(N)	0.8	0.80033	18	25	43	24.66	0.1	0.25	0.73380	0.05	0.04802	0.19967	2	7

Single Stage	A regular, no-interim look design. It is provided for reference.
Min N	The design with the smallest N. Simon called this the "Minimax" design.
Min E(N)	The design with the smallest E(N). Simon called this the "Optimum" design.
Admissible	A compromise design developed using a Bayesian argument by Jung (2004). It has properties between Simon's Minimax and Optimum designs.
α, β met	A design that meets both the Type-I and Type-II error requirements but is not optimal or admissible.
Power	The probability of rejecting the null hypothesis that the proportion responding (P) to the treatment is less than or equal to P0 when this hypothesis is false (when P = P1). Power = Pr(rejecting P ≤ P0 | P ≥ P1).
Target Power	The power that is required.
Actual Power	The power that is achieved when the actual response proportion is P1.

N1	The sample size in the first stage.
N2	The sample size in the second stage.
N	The combined (total) sample size of both stages.
E(N)	The expected (average) sample size if this design were repeated many times. It is calculated assuming P = P0.
P0	The maximum response proportion that will be considered as POOR (Not worth pursuing).
P1	The minimum response proportion that will still be considered GOOD (Worth pursuing).
PET	The probability of early termination of the study. It is the probability that the design will be stopped after stage 1. It is calculated assuming P = P0.
Alpha	The probability of rejecting that P ≤ P0 when this is true.
Target Alpha	The alpha that is required.
Actual Alpha	The alpha that is achieved.
Actual Beta	The beta (Type-II) error rate that is required.
R1	The critical value of the first stage. Stop trial if response ≤ R1. Continue trial if response is > R1.
R2	The critical value after both stages. Reject H0: P ≤ P0 if response > R.

Summary Statements

A two-stage phase II single-arm clinical trial design (Simon, 1989) will be used to test whether the proportion responding (P) warrants continuation to the next phase (H0: P ≤ 0.1 versus H1: P ≥ 0.25). For the design that minimizes the expected sample size, with a Type I error rate of 0.05, a power of 80%, and a good treatment minimum response rate of 0.25, the total number of subjects required if the study continues to the second stage is 43. In the first stage, 18 subjects will be needed, with an additional 25 subjects in the second stage, if necessary. The expected (average) sample size of this design is 24.66, with a probability of stopping after Stage 1 of 0.7338. With 18 subjects at the first stage, the trial should be discontinued if 2 or fewer respond to the treatment. If the trial continues to the second stage, the treatment efficacy is rejected if 7 or fewer of the total 43 subjects respond to the treatment. Otherwise, if the number that responds is greater than 7, H0 is rejected in favor of continuance to the next phase.

References

Simon, Richard. 'Optimal Two-Stage Designs for Phase II Clinical Trials', Controlled Clinical Trials, 1989, Volume 10, pages 1-10.

Jung, S.H., Lee, T., Kim, K.M., and George, S.L. 'Admissible two-stage designs for phase II cancer clinical trials', Statistics in Medicine, 2004, Volume 23, pages 561-569.

This report shows three optimal designs after showing the single stage design. The first is the Min N (minimax) solution. This is the design with the smallest total sample size (*N*). The last is the Min E(N) (optimum) design—the one that minimizes the average sample size. Two admissible designs are also displayed.

Example 2 – Validation using Simon (1989)

Simon (1989) page 4 in his Table 1 presents designs for several scenarios. The first row of the table sets *P0* to 0.05, *P1* to 0.25, alpha to 0.10, and power to 0.90.

The minimax, Min N, design is *N1* = 13, *R1* = 0, *N* = 20, and *R* = 2.

The optimal, Min E(N), design is *N1* = 9, *R1* = 0, *N* = 24, and *R* = 2.

We will now run this example through **PASS**.

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

Search Type	**Automatic**
Power (Probability of Rejecting H0 P ≤ P0 | P = P1)	**0.90**
Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0)	**0.1**
P0 (Maximum Response Rate of a Poor Treatment)	**0.05**
P1 (Minimum Response Rate of a Good Treatment)	**0.25**
Multiplier for Upper Search Limit of N	**1.25**

Output

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

Designs

Solve For:	Sample Size
Beta Target:	0.1
N Search Limit Multiplier:	1.25

	Power		Sample Size				Response Rate			Alpha			Critical Value	
Design Type	Target	Actual	N1	N2	N	E(N)	Maximum of a Poor Treatment P0	Minimum of a Good Treatment P1	Probability of Early Termination PET	Target	Actual	Actual Beta	R1	R2
Single Stage	0.9	0.90874	20		20		0.05	0.25		0.1	0.07548	0.09126	2	
Min N	0.9	0.90295	13	7	20	16.41	0.05	0.25	0.51334	0.1	0.07356	0.09705	0	2
Admissible	0.9	0.90544	11	10	21	15.31	0.05	0.25	0.56880	0.1	0.07837	0.09456	0	2
Admissible	0.9	0.90504	10	12	22	14.82	0.05	0.25	0.59874	0.1	0.08311	0.09496	0	2
Min E(N)	0.9	0.90284	9	15	24	14.55	0.05	0.25	0.63025	0.1	0.09313	0.09716	0	2

PASS calculates the same Min N and Min E(N) designs.

Example 3 – Using Custom Search

This example will show the use of the custom search option. In this example, we will find that the two optimal Simon designs are not of practical use. We will use the custom search option to find a more appealing design.

Suppose a design is wanted for the case alpha = 0.05, power = 0.80, *P0* = 0.7, and *P1* = 0.9. 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 3a** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Search Type	**Automatic**
Power (Probability of Rejecting H0 P ≤ P0 | P = P1)	**0.8**
Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0)	**0.050**
P0 (Maximum Response Rate of a Poor Treatment)	**0.7**
P1 (Minimum Response Rate of a Good Treatment)	**0.9**
Multiplier for Upper Search Limit of N	**1.25**

Output

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

Designs

Solve For:	Sample Size
Beta Target:	0.2
N Search Limit Multiplier:	1.25

							Response Rate							
	Power		Sample Size				Maximum of a Poor Treatment	Minimum of a Good Treatment	Probability of Early Termination	Alpha		Actual	Critical Value	
Design Type	Target	Actual	N1	N2	N	E(N)	P0	P1	PET	Target	Actual	Beta	R1	R2
Single Stage	0.8	0.85789	28		28		0.7	0.9		0.05	0.04743	0.14211	23	
Min N	0.8	0.80096	23	3	26	23.16	0.7	0.9	0.94616	0.05	0.04526	0.19904	19	21
Min E(N)	0.8	0.80418	6	21	27	14.82	0.7	0.9	0.57983	0.05	0.04924	0.19582	4	22

Notice that neither of the two optimal designs is of practical use. The Min N (or minimax) design sets N1 at 23 and N at 26. Thus, this design does not call for the interim analysis until 88% of the subjects are observed. Not many researchers would adopt such a design.

The Min E(N) (or optimum) design is also disappointing for the opposite reason. In this design, N1 is 6 and N is 27. Thus, only 22% of the subjects are observed before the interim analysis is planned. Not many researchers would adopt this design either.

And so, we are stuck with two designs that both have obvious flaws.

We will now conduct a search of near optimal designs that have more favorable properties. This is done as follows.

Suppose a design is wanted for the case alpha = 0.05, power = 0.80, *P0* = 0.7, and *P1* = 0.9. 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 procedure window. You may then make the appropriate entries as listed below, or open **Example 3b** by going to the **File** menu and choosing **Open Example Template**.

Design Tab

Search Type	**Custom**
Power (Probability of Rejecting H0 P ≤ P0 | P = P1)	**0.8**
Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0)	**0.050**
P0 (Maximum Response Rate of a Poor Treatment)	**0.7**
P1 (Minimum Response Rate of a Good Treatment)	**0.9**
N (Combined Sample Size) Min	**26**
N (Combined Sample Size) Max	**27**
R (Rejection Number) Min	**4**
R (Rejection Number) Max	**27**
N1 (First Stage Sample Size) Min	**12**
N1 (First Stage Sample Size) Max	**15**
R1 (First Stage Rejection Number) Min	**2**
R1 (First Stage Rejection Number) Max	**14**

Output

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

Designs

Solve For:	Sample Size
Beta Target:	0.2
Search Parameter Ranges:	N (26 to 27), R (4 to 27), N1 (12 to 15), R1 (2 to 14)

	Power		Sample Size				Response Rate			Alpha			Critical Value	
Design Type	Target	Actual	N1	N2	N	E(N)	Maximum of a Poor Treatment P0	Minimum of a Good Treatment P1	Probability of Early Termination PET	Target	Actual	Actual Beta	R1	R2
Single Stage	0.8	0.85789	28		28		0.7	0.9		0.05	0.04743	0.14211	23	
Min N & E(N)	0.8	0.82226	12	15	27	15.79	0.7	0.9	0.74718	0.05	0.04955	0.17774	9	22
α, β met	0.8	0.80881	13	14	27	15.83	0.7	0.9	0.79752	0.05	0.04716	0.19119	10	22

Now we have two designs to choose from. Both designs have similar total sample sizes of 27. The Min E(N) design had an E(N) of 14.82. These designs only increase the expected sample size by one. However, now N1 is 12 or 13 which translates to an interim analysis at about 50% of the total sample. We would recommend the second since it is nearly balanced.