Chapter 130

Three-Stage Phase II Clinical Trials

Introduction

Phase II clinical trials determine whether a drug or regimen has sufficient activity against disease to warrant more extensive study and development. In a three-stage design, the patients are divided into three 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. A similar interim analysis is conducted at the end of the second stage.

This module finds designs that meet the error rate (alpha and beta) criterion and minimize the expected sample size. The formulation is given in Chen (1997). Extending Chen's work, our algorithm allows the investigation of near-optimal designs that may have other useful properties.

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.

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 it is true.

A three-stage phase II design can be represented by six numbers: *R1*, *N1*, *R2, N2, R3* and *N3*. *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. *N2* is the total sample size of stages one and two. *R2* is the critical value in the second stage. If *R2* or fewer responses occur in the *N2* patients, the drug is rejected. *N3* is the combined sample size of all three stages. *R3* is the critical value in the combined sample. If *R3* or fewer of the *N3* patients respond, the drug is rejected.

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

$$E(N_E) = N1 + (1 - PET1)(N2 - N1) + (1 - PET2)(N3 - N2)$$

where *PET1* is the probability of early termination of the study after stage one and *PET2* is the probability of early termination after stage two.

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

$$\Pr(\text{reject}|P, N1, R1, R2, N2, R3, N3) = PET1 + PET2 + PET3$$

where

$$PET1 = B(R1|P, N1)$$

$$PET2 = \sum_{X1=R1+1}^{\min(N1,R2)} b(X1|P, N1)B(R2 - X1|P, N2 - N1)$$

$$PET3 = \sum_{X1=R1+1}^{\min(N1,R3)} b(X1|P, N1) \sum_{X2=R2+1-X1}^{\min(N3-N2,R3-X1)} b(X2|P, N2 - N1)B(R3 - X1 - X2|P, N3 - N2)$$

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

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

The two error rate constraints are

$$\Pr(\text{reject}|PO, N1, R1, R2, N2, R3, N3) \geq 1 - \alpha$$

and

$$\Pr(\text{reject}|P1, N1, R1, R2, N2, R3, N3) \geq \beta$$

Optimum Design

The optimum design minimizes the average sample size, *E*(*N*), while meeting the error rate constraints. This design is found through an exhaustive search of all possible designs. This search may take several minutes to complete.

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 *N1* = 5, *N2* = 25, and *N3* = 26. This design is poor because the bulk of the subjects are tested in the second phase. Most researchers would rather have more balance in the sample sizes of the three stages. For reasons like this, the actual optimal design may be replaced by another, sub-optimal, design.

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 when *P* = *P0*, a proportion α of them will be falsely terminated due to chance occurrences.

The point is that 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 researcher must also recognize that when the sampling plan is changed, the exact probability statements can no longer be calculated.

Custom Search - Mininum N (Combined Sample Size)

N is the combined sample size of the three stages of the design. This parameter sets the minimum value of *N3* that is used during the search. The optimum value of *N3* 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 three-stage sample size will be less MIN, this provides a reasonable starting point for a search for *N*. You can also enter a value like MIN-x where *x* is a positive integer. This will cause the search to begin *x* units below the MIN.

Example 1 – Calculating the Power and Validation using Chen (1997)

Chen (1997) provides the minimax and optimum design for the case Alpha = 0.05, Beta = 0.20, *P0* = 0.05, and *P1* = 0.25. The optimum design is 0/8, 1/13, and 2/19. The minimax design is 0/12, 1/15, and 2/16.

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

Designs to Display	**Optimum designs only**
Power	**0.80**
Alpha	**0.05**
P0 (Poor)	**0.05**
P1 (Good)	**0.25**
N Min	**Min**
N Max	**Best 2**

Note that the search may take several minutes to run, depending on the speed of your computer.

Output

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

Numeric Results

	Power		Reject Treatment if Response Rate			Expected Sample Size	Response Rate		Probability of Early Termination		Alpha		Beta	
Design Type	Target	Actual	Stage 1 = R1/N1	Stage 2 ≤ R2/N2	Stage 3 ≤ R3/N3	E(N)	Poor P0	Good P1	Stage 1 Pet1	Stage 2 Pet2	Target	Actual	Target	Actual
Single Stage	0.8	0.803	2/16			16.00	0.05	0.25			0.05	0.043	0.2	0.197
Minimax	0.8	0.801	0/12	1/15	2/16	13.55	0.05	0.25	0.540	0.833	0.05	0.043	0.2	0.199
Optimum	0.8	0.805	0/8	1/13	2/19	10.41	0.05	0.25	0.663	0.880	0.05	0.049	0.2	0.195

Design Type	This column lists the design constraints that are obeyed by the design analyzed on this row.
Power	This is the probability of rejecting the hypothesis that the proportion responding to the treatment is less than or equal to P0 when this hypothesis is false. That is, Power = Pr(rejecting P ≤ P0 | P ≥ P1).
Target Power	The power that was desired.
Actual Power	The power that was achieved.
R1	The drug rejection number in the first stage.
N1	The sample size in the first stage.
R2	The drug rejection number in the second stage.
N2	The sample size in the first and second stages.
R3	The drug rejection number in the third stage.
N3	The combined sample size of all three stages.
E(N)	The average sample size if this design is repeated many times.
P0	The response proportion of a poor drug.
P1	The response proportion of a good drug.
Pet1	The probability of early termination at the first stage.
Pet2	The probability of early termination at the second stage.

Alpha	The probability of rejecting that P ≤ P0 when this is true. This is often called the Type-I error rate.
Target Alpha	The alpha that was desired.
Actual Alpha	The alpha that was achieved.
Beta	The probability of rejecting that P ≥ P1 when this is true. This is often called the Type-II error rate.
Target Beta	The beta that was desired.
Actual Beta	The beta that was achieved.

Summary Statements

A three-stage phase II single-arm clinical 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 ≥ 0.25). For this design (that seeks to minimize 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 third stage is 19. In the first stage, 8 subjects will be needed, with a total of 13 subjects in the second stage, and a total of 19 in the third stage, if necessary. The expected (average) sample size of this design is 10.41, with a probability of stopping after Stage 1 of 0.663, and a probability of stopping after Stage 2 of 0.88. With 8 subjects at the first stage, the trial should be discontinued if 0 or fewer respond to the treatment. If the trial continues to the second stage, with 13 total subjects at this stage, the trial is stopped if 1 or fewer respond. If the third stage is reached, the treatment efficacy is rejected if 2 or fewer of the total 19 subjects respond to the treatment. Otherwise, if the number that responds is greater than 2, H0 is rejected in favor of continuance to the next phase.

References

Chen, T. T. 1997. 'Optimal Three-Stage Designs for Phase II Cancer Clinical Trials.' Statistics in Medicine, Volume 16, pages 2701-2711.

This report shows three designs. The first is the smallest single stage design. The second is the minimax solution. This is the design with the smallest total sample size (*N*). The third is the optimum design—the one that minimizes the average sample size.

Note that **PASS** matches the results of Chen (1997).