BIOSTATISTICAL METHODS

FOR TRANSLATIONAL & CLINICAL RESEARCH

Two-stage Designs:

EARLY-PHASE CLINICAL TRIALS

This lecture covers a very special form of phase II clinical trials: two-stage design.

A small group of patients are enrolled in the first stage; the enrollment of another group of patients in stage 2 is "conditional" on the outcome of the first group.

The activation of the second stage depends on an adequate number of responses observed from the first stage.

Rationale:

Why two stages? Do not want to enroll a large group of patients (in conventional one-stage designs) when not sure if the treatment is effective. If treatment is not effective, cancer would kill the patients. In two-stage designs, the second stage is not activated if the first group/stage shows that the treatment is not effective

Rationale:

If treatment is not effective, enrolled patients might die because of the disease.

What lost is “opportunity”; they could survive with a better treatment from another trial.

There are more than one method – some are recent, but the emphasis of this lecture is on a very popular method called “two-stage Simon’s Design”.

This design uses a “computer search” to meet certain optimal requirement; it does require some special program; research organizations and health centers have this software.

Phase I trials provide information about the MTD; it is important because most cancer treatments must be delivered at maximum dose for maximum effect.

Patients may die from toxicity or side effects and, if not treated "enough", they might die from the disease too. Phase I trials provide little or no information about efficacy; patients are diverse with regard to their cancer diagnosis and are treated at different doses - only 3 or 6 at a dose – even one at a dose by fast-track design.

A phase II trial of a cancer treatment is an uncontrolled trial (most trials of phase II are one-arm, open-label) to obtain an estimate of the "degree of anti-tumor effect". The proportion of patients who "tumors shrink by at least 50% which lasts for at least 4 weeks" is often the primary endpoint. The aim is to see if the agent has sufficient activity against a specific type of tumor to warrant its further development (to combine with other drugs in a phase III trial comparing survival results with a standard treatment).

It is desirable to find out about the anti-tumor capacity of new agents and to determine if a treatment is sufficiently promising to warrant a major controlled evaluation. However, recall that there are three basic objectives in conducting phase II clinical trials:

(1) Benefit the patients

(2) Screen agent/drug for anti-tumor activity

(3) Extend knowledge of toxicology and pharmacology of drug/agent.

The first aim is to benefit the patients

The problem is that, if the agent has no or low anti-tumor activity, patients in the phase II trial might die from the disease. Therefore, we often wish to minimize the number of patients treated with an ineffective drug.

Early acceptance of an highly effective drug is permitted but very rare in phase II trials; however, it is ethically imperative to exercise early termination when the drug has no or low anti-tumor activity.

GEHAN'S TWO-STAGE DESIGN

- The first and most commonly used design for many years was developed by Gehan (1961); this design has been popular- more so in the 70's.
- The same Gehan who invented the "generalized Wilcoxon test" (two years later).
- It has two stages; the primary aim Gehan's design is to estimate the response rate - two-stage feature is an option for "screening" of agents worthy of further development.

GEHAN'S DESIGN

- The first stage enrolls 14 patients; if no responses are observed, trial is terminated;
- If at least one response is observed among the first 14 patients of stage 1, the second stage of accrual is activated in order to obtain an estimate of the response probability having a pre-specified standard error (SE).
- Patients from both stages are used in the estimation of the response rate.

RATIONALE FOR EARLY TERMINATION

The probability of observing no responses among 14 patients is less than .05 if the response probability is greater than 20%

$$(.20)^0(.80)^{14} = .044$$

$$\pi^0(1-\pi)^{14} \leq .044 \text{ if } \pi \geq .20$$

Implicitly, response rates over 20% are considered promising for further studies.

If no responses are observed in the first stage of 14 patients, trial is terminated. It is stopped because we can conclude that $\pi<.2$ or 20%, not worthy of further investigation.

SECOND STAGE

The number of patients n_2 accrued in the second stage depends on the number of responses observed in the first stage (because patients from both stages are used in the estimation of the response rate) and the pre-determined standard error;

$$SE(p) = \sqrt{\frac{\pi(1-\pi)}{14+n_2}}$$

However, Gehan's design is often used with a second stage of $n_2 = 11$ patients. This accrual provides for estimation with no more than 10% standard error (SE≤.10)

$$SE(p) = \sqrt{\frac{\pi(1-\pi)}{25}} \leq \sqrt{\frac{(.50)(.50)}{25}} = .10$$

CRITIQUES

- The size of the first stage is “fixed”; it may not be optimal for the underlying aim of early termination “if the drug has no or low anti-tumor activity”.
- It serves investigators & drug companies more - not the patients enrolled in the trial.
- With a standard error of 10%, it corresponds to a very broad 95% confidence interval; reducing SE to, say, 5% would lead to a sample size too large for phase II trials (but this is an universal problem for phase II trials- not just Gehan’s).

AN MAJOR WEAKNESS

The more serious problem is in the first critique.

Gehan's design provides an option for screening of agents worthy of further development; those with response rates of 20% or more. However, it does not help to achieve the aim of early termination when the drug has no or low anti-tumor activity, a very important ethical concern

For example, even a poor drug with a true response probability of 5%, there is a 51% chance of obtaining at least one response in the first 14 patients and, therefore, activating the second stage accrual.

$$\Pr(\text{at least 1 response}) = 1 - (.05)^0(.95)^{14} = .51$$

That is, there might be high probabilities to enroll more patients (in the second stage) to be treated by an inefficient drug.

Gehan's two-stage design is being used in some trials because of its simplicity.

TWO-STAGE SIMON'S DESIGN

In phase II trials, the ethical imperative for early termination occurs when the drug has low anti-tumor activity; the "Two-stage Simon's Design" is currently a popular tool to achieve that.

The trial is conducted in two stages with the option to stop the trial after the first or after the second stage (and not recommending the agent for further development).

The basic approach is to "minimize" expected sample size when the true response is low - say, less than some pre-determined uninterested level.

PROBLEM'S STATISTICAL SETUP

- **Endpoint:** (binary) Tumor Response: yes/no
- **Null Hypothesis:** H_0: $\pi = \pi_0$; π is the true response (say, proportion of patients whose tumors shrink by at least 50%) and π_0 is a pre-determined uninterested/undesirable level.
- **Alternative Hypothesis:** H_A: $\pi = \pi_A$; π_A is some desirable level that warrant further development.
- **Type I and type II errors:** α and β
- **Basis for decision:** minimize the number of patients treated in the trial if H_0 is true.

The response rate under the Null Hypothesis π_0 is considered as “dangerous” because patients might die from the disease. It could be, say, the placebo effect. The Alternative Hypothesis, π_A, is the investigators hypothesized value – formed from preclinical results and some from the completed Phase I Trial.

THE DESIGN

- Two stages (the design could have more than two stages, but less practical and not used)
- Enroll n_1 patients in stage 1; the trial is stopped if r_1 or fewer responses are observed, goes on to the second stage otherwise.
- Enroll n_2 patients in stage 2; the trial is not recommended for further development if a total of r ($r = r_1 + r_2$, r_2 from second stage) or fewer responses are observed in both stages.

Our job in designing the trial is to find the numbers n_1, n_2, r_1, and r_2 so as to minimize the number of patients treated in the trial if H_0 is true. The level of Type I errors and Statistical power are pre-determined.

DRAWBACK

The main practical consideration is that evaluation of a patient's response is usually not instantaneous and may require observations for weeks or months. Consequently, patient accrual at the end of stage 1 may have to be suspended until it is determined whether the criterion for continuing is satisfied. Such suspension of accrual is awkward for physicians who are contributing patients to the study; and is the main reason for not considering more than two stages.

PET: PROB OF EARLY TERMINATION

$$\mathbf{PET}(\pi) = \mathbf{B}(r_1; n_1, \pi)$$

where B(.) denotes the cumulative Binomial probabilit y, and π is the true response.

That's the probability to have r_1 responses or fewer in first stage.

EXPECTED SAMPLE SIZE

$$EN(\pi) = n_1 + [1 - PET(\pi)] * n_2$$

Both PET(π), the probability of early termination, and EN(π), expected sample size, are function of the response rate π.

DECISION NOT TO RECOMMEND

- The drug may not be recommended either after 1 or 2 stages; the probability is PNC(π).
- We will terminate the trial at the end of the first stage and not recommending the drug if r_1 or fewer responses are observed, or
- We will not recommend the drug at the end of the second stage if r ($r = r_1 + r_2$) or fewer responses are observed; some of the responses after the first (r_1) may come from stage 1, some from stage 2.

PNC: PROB OF NOT RECOMMENDING

The drug is not recommended if the trial is terminated early (i.e. fewer than r_1 responses are observed in the first stage) OR fewer than $r = r_1 + r_2$ are observed; some of the responses (say, x) may come from stage 1 and some from stage 2 (say, r-x).

$$PNC(\pi) = B(r_1; n_1, \pi) + \sum_{x=r_1+1}^{r} b(x; n_1, \pi) B(r - x; n_2, \pi)$$

The probability of not recommending, PCN(π), is also a function of the response rate π.

TWO TYPES OF ERRORS

$$(\text{Type I errors}) : \alpha = 1 - PNC(\pi_0)$$

$$(\text{Type II errors}) : \beta = PNC(\pi_A)$$

SIMON'S APPROACH

The design approach considered by Simon is to specify the parameters π_0, π_1, α, and β; then determine the two-stage design that satisfies the errors probabilities α and β and minimizes the expected sample size EN when the response probability is π_0; i.e. minimizing EN(π_0).

DETAILED IMPLEMENTATION

It's a search with the help of a computer program.

Step #1: for each integer n and n_1, in the range (1,n-1), determine the integers r_1 and r which satisfies the error constraints and minimizes $EN(\pi_0)$.

(i) This is found by searching over the range r_1 in $(0,n_1)$; for each value of r_1, determine the value of r satisfying the type II error rate; then

(ii) To see whether the set of parameters (n,n_1,r_1,r) satisfied type I error rate; if it did, to compare the expected sample size $EN(\pi_0)$ to the one achieved previously, and continue to search with a different r_1.

Step #2a: Keeping n fixed, search over the range of n_1,(1,n-1), repeating the process in step1

Step #2b: search over the range of n up to, say, 50, as commonly used in phase II trials - repeating the process in step 1.

Simon called this "**Optimal Design**", in the sense to achieve the objective of early termination when the drug has no or low anti-tumor activity (key step: check EN(π_0))

There is an option where, in step #1, we only check against the constraints imposed by the two types of errors but skip checking for EN(π_0); then keep the same steps 2a and 2b. The first two-stage design found would satisfy the error constraints and has smallest total maximum/potential sample size n.

Simon called this the "Minimax Design"

Usually, the minimax two-stage design has the same maximum sample size n as the smallest single-stage design that satisfies the error probabilities. However, because of the early termination option (after first stage), the minimax two-stage design has a smaller expected sample size under H_0.

EXAMPLE 1A: OPTIMAL

- Let consider: π_0 =.10 and π_1 = .30
- and take: α = .05 and β = .20
- Stage 1: reject the drug if response ≤1/11 (enroll 11 and terminate if 0 or 1 response)
- Stage 2: (2 or more responses from stage 1): reject the drug if response ≤ 5/29 (enroll 18 in the 2nd stage; reject if total responses ≤ 5)
- Expected sample size EN(π = .10) = 15.0
- PET(π = .10)= .74

EXAMPLE 1B: MINMAX

- Same design parameters: π_0 =.10 and π_1 = .30
- and same error rates: α = .05 and β = .20
- Stage 1: reject the drug if response ≤1/15 (enroll 15 and terminate if 0 or 1 response)
- Stage 2: (2 or more responses from stage 1): reject the drug if response ≤ 5/25 (enroll 10 in the 2nd stage; reject if total responses ≤ 5)
- Expected sample size EN(π = .10) = 19.5
- PET(π = .10)= .55

Optimal Design:

Rejection: if

$\leq$ 1/11 or $\leq$ 5/29

EN(p=.10) = 15.0

PET(p=.10)= .74

Minimax Design:

Rejection: if

$\leq$ 1/15 or $\leq$ 5/25

EN(p=.10) = 19.5

PET(p=.10)= .55

Response Rates: π_0 =.10 and π_1 = .30

Error rates: α = .05 and β = .20

If "one-stage", we would need n=32

Minimax design enrolls no more than 25 patients (versus 29 for optimal); however, it (always) enrolls more subjects in the first stage (15 versus 11), (always) has a lower probability for early termination if the drug performs poorly (here, .55 versus .74), therefore, it has larger expected sample size (than optimal design, 19.5 versus 15).

In some cases, the “minimax” design may be more attractive than the “optimal” design. This is the case where the difference in expected sample sizes EN(π_0) is small and the patient accrual rate is low/slow. We would always need to check out both!

EXAMPLE 2A: OPTIMAL

- Let consider: π_0 =.10 and π_1 = .30
- and take: α = .10 and β = .10
- Stage 1: reject the drug if response ≤1/12 (enroll 12 and terminate if 0 or 1 response)
- Stage 2: (2 or more responses from stage 1): reject the drug if response ≤ 5/35 (enroll 23 in the 2nd stage; reject if total responses ≤ 5)
- Expected sample size EN(π = .10) = 19.8
- PET(π = .10)= .65

EXAMPLE 2B: MINMAX

- Same design parameters: π_0 =.10 and π_1 = .30
- and same error rates: α = .10 and β = .10
- Stage 1: reject the drug if response ≤1/16 (enroll 16 and terminate if 0 or 1 response)
- Stage 2: (2 or more responses from stage 1): reject the drug if response ≤ 4/25 (enroll 9 in the 2nd stage; reject if total responses ≤ 4)
- Expected sample size EN(π = .10) = 20.4
- PET(π = .10)= .51

Optimal Design:

Rejection: if

$\leq$ 1/12 or $\leq$ 5/35

EN(p=.10) = 19.8
PET(p=.10)= .65

Minimax Design:

Rejection: if

$\leq$ 1/16 or $\leq$ 4/25

EN(p=.10) = 20.4
PET(p=.10)= .51

Response Rates: π_0 =.10 and π_1 = .30

Error rates: α = .10 and β = .10

If "one-stage", we would need n=35

It is still true that minimax design enrolls no more than 25 patients (versus 35), more subjects in the first stage (16 versus 12), lower probability for early termination if the drug performs poorly (.51 versus .65), and larger expected sample size (20.4 versus 19.6). But the reduction in expected sample size is negligible; on the other hand, if the accrual is slow - say, 10 per year - it could take a year longer to complete the optimal design than the minimax design.

Suggested Readings:

Search and find (to keep) the papers:
(1) Simon, Controlled Clinical Trials, 1989
(2) Duffy and Santner, Biometrics, 1987 (for Confidence intervals)