STUDY DESIGNS

IN BIOMEDICAL RESEARCH

PHASE I CLINICAL TRIALS

Steps to New Drug Discovery

Get idea for drug target

↓

Develop a bioassay

↓

Screen chemical compounds in assay

↓

Establish effective and toxic amounts

↓

File for approval as an Investigational New Drug (IND)

(After an IND is applied, it's the starting point of clinical trials)

PHASES OF CLINICAL TRIALS

- Phase I: First human trial to focus on safety
- Phase II: Small trial to evaluate efficacy
- Phase III: Large controlled trial to demonstrate efficacy prior to FDA approval
- Phase IV: Optional, post-regulatory approval, to provide the medicine's more comprehensive safety and efficacy profile

Phase I and Phase II are referred to as "early-phase clinical trials"

When designing cancer clinical trials for development and evaluation of therapeutic interventions, two special aspects must be taken into consideration: (1) the target population (cancer patients) and (2) the fact that all anti-cancer drugs under investigation are cytotoxic agents.

It is true that newer drugs are safer – and more efficient, but toxicity is still a serious concern for everyone involved in clinical research.

Potential therapeutic agents for cancer treatment can induce severe safety concern - even at lower dose levels; they can generate severe toxicities than most of pharmaceutical agents for treatment of other diseases. Possible adverse effects may be irreversible, even fatal. As a result, phase I trials are conducted only on cancer patients - not healthy volunteers.

Patients in cancer clinical trials are those with malignant tumors; most cancers are life-threatening and the disease process is usually irreversible. Patients in phase I trials are mostly terminal cancer patients who have failed all standard therapies and for whom the new anti-tumor agent being tested may be the last hope

At this stage, the investigator is facing a classic, fundamental dilemma: it's a conflict of scientific versus ethical intent.

We put some patients at risk for their own benefits and benefits of others; and it's an unavoidable conflict because these patients have failed under all standard therapies. We need to reconcile the risks of toxicity to patients with the potential benefit to these same patients and make an efficient design to use no more patients than necessary.

The primary scientific objective of the evaluation of new chemotherapeutic agents in cancer patients in phase I trials is to employ an efficient dose-finding design to reach “the maximum dose with an acceptable and manageable safety profile” for use in subsequent phase II trials. The most commonly used design is the “standard design”, or some of its variations.

Phase I and II clinical trials present special difficulties because they involve use of agents whose spectrum of toxicity and likelihood of benefits are poorly understood/defined. There were "pre-clinical" studies – e.g. In Vivo & In Vitro experiments and bioassays – but the subjects were animals (In Vivo) or human tissues (In Vitro). And inferences across species are never easy, nor precise.

In term of “popularity”, there are many more phase I and phase II trials than phase III trials (& phase IV trials are more rare because they “optional”). That is because out of many phase I trials, perhaps 10 or so, may be only one agent is found suitable to go on to phase II. And out of many phase II trials, again perhaps 10 or so, may be only one agent is found good enough to proceed to a phase III trial. In more recent time, phase II trials more often involve drug combinations

CANCER PHASE I TRIALS

- Different from other phase I clinical trials, phase I clinical trials in cancer have several main features.
- The efficacy of chemotherapy is associated with a non-negligible risk of severe toxic effect, often fatal, so that ethically, such drugs can be investigated only in cancer patients; and only a small number of patients are available for the trial- any trial in any phase.
- These patients are at very high risk of death in the short term under all standard therapies. At low doses little or no efficacy is expected.
- A slow intra-patient dose escalation is not practically possible.

SETTING & GOAL OF PHASE I CANCER TRIALS

- Patients from standard treatment failure
- New Drug: No efficacy at low doses, will have toxicity at high doses- maybe severe, fatal
- Dose range: little known (first in human)
- Goal: Maximum Tolerated Dose (MTD), reasonable efficacy & tolerable toxicity.

STANDARD DESIGN

Starting Dose:

The starting dose selection of a phase I trial depends heavily on pharmacology and toxicology from pre-clinical studies. Although the translation from animal to human is not always a perfect correlation; generally, toxicity as a function of body weight is assumed roughly constant across species.

From the toxicity and body weight relationship, it is also often implicitly considered that mouse LD10 is about the same as human MTD. However, the use of a second species has also shown to be necessary because in approximately 90 reviewed drugs, mouse data alone was insufficient to safely predict the human MTD (Arbuck, 1996).

For most investigators, the first dose is often chosen at about one-tenth of the mouse LD10 (at which 10% of mice die). However, some are more cautious & proposed to use the smaller of:

(i) One tenth of the mouse LD10, and

(ii)One third of the beagle dog LD10.

Number of Doses:

Phase I trials are designed with 3 to 8 doses; most with 5 doses – some with even 4 doses.

Spacing Between Doses:

Generally, the dose levels are selected in order that the percentage increments between successive doses diminish as the dose is increased; for example, (i) equally-spaced on the log scale or (ii) a modified Fibonacci sequence is often employed (increases of 100%, 67%, 50%, 40%, then 33% for subsequent doses if more than 5 are planned); this follows a diminishing pattern, with modest increases

DOSE ESCALATION

- Standard Design is sometimes referred to as "3-and-3 Design". Start at lowest dose, enroll groups of 3 pts

(i) Move up if none of first 3 have toxicity;

(ii) If two or three patients have dose-limiting toxicity (DLT), stop.

(iii) If one of 3 has toxicity, enroll 3 more at same dose and move up if none of the second cohort have toxicity. Otherwise, stop

- Moving up, repeat the process at new dose

MAXIMUM TOLERATED DOSE

- If dose escalation not possible, i.e. being stop at a given dose, this dose is considered as "above MTD" – have to go down to find MTD
- When a dose is judged as above the MTD, the next lower dose is (often) declared the MTD in the designs without dose de-escalation (it's not completely universal; some investigators only step down half a dose).

Design Characteristics:

If r is the toxicity rate of the current dose, the probability of escalating after only 3 patients is:

$UP_3 = (1-r)^3$

The probability of stopping after only 3 patients is:

$STOP_3 = 3r^2(1-r) + r^3$

And probability to stop (3 or 6 patients):

$STOP_{3/6} = [3r^2(1-r) + r^3] + [3r(1-r)^2][1-(1-r)^3]$

The probability that the second cohort of needed is:

$NEED_6 = 1- (UP_3 + STOP_3)$

Rate, r	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70
$STOP_{3/6}$	0.03	0.09	0.29	0.51	0.69	0.83	0.92	0.97
UP_3	0.86	0.73	0.51	0.34	0.22	0.13	0.06	0.03
$STOP_3$	0.01	0.03	0.10	0.22	0.35	0.50	0.65	0.78
$NEED_6$	0.13	0.24	0.39	0.44	0.43	0.37	0.29	0.19

(1) The cohort is limited to 3 pts each with ≥56% probability: It requires few patients

(2) For more extreme DLT probabilities (5% & 70%), the cohort is only expanded to 6 patients with probability of less than 20%

(3) When the rate is 30% or higher, the probability to stop is ≥51%: It's rather safe

Rate, r	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70
$STOP_{3/6}$	0.03	0.09	0.29	0.51	0.69	0.83	0.92	0.97
UP_3	0.86	0.73	0.51	0.34	0.22	0.13	0.06	0.03
$STOP_3$	0.01	0.03	0.10	0.22	0.35	0.50	0.65	0.78
$NEED_6$	0.13	0.24	0.39	0.44	0.43	0.37	0.29	0.19

(1) If 2 or 3 pts show DLT ($STOP_3$), we are 90% sure that r≥20%

(2) If no patients show DLT (UP_3), we are 90% sure that r≤55%

(3) If r=10%, 91% chance to escalate; if r=60%, 92% chance to stop

Brief conclusions are:

(1) Standard Design is rather safe; and

(2) Standard Design would not require too many patients. However, “details” are not there … yet!

A FEW RESEARCH QUESTIONS

Here are 3 basic more three research questions:

(1) What is the actual expected toxicity rate of the MTD selected by the standard design,

(2) Is the expected toxicity rate of the selected MTD by the standard design robust, and

(3) Can improvement be made to the standard design so that we can get to the next phase quicker but still safe enough? What are alternatives to the Standard Design which could help to accelerate the process?

Some important parameters, such as expected toxicity rate and trial size can be estimated/approximated

If r_i is the toxicity rate of the dose i, the probability of stopping at dose i, after 3 or 6 patients is :

$$STOP_{3/6} = [3r^2(1-r) + r^3] + [3r(1-r)^2][1-(1-r)^3]$$

$$= p_{[i]}$$

That is the probability to stop at dose i, given that you already at dose i

If r_i is the toxicity rate of the dose i and $p_{[i]}$ the probability of stopping at dose i, the probability of escalating to dose i,

$p_{(1)} = 1$

$p_{(2)} = 1 - p_{[1]}$

$p_{(i)} = p_{(i-1)}[1 - p_{[i-1]}]$ for $i = 2,3,\ldots$

Then we have:

$$\text{Expected Toxicity Rate of MTD} = \sum_i r_i p_{[i+1]}$$

$$\text{Expected Trial Size} = \sum_i p_{(i)}[(3)(UP_3 + STOP_3) + (6)(NEED_6)]$$

(Recall: $p_{[i+1]}$, the probability of stopping at dose (i+1), is also the probability that dose i is selected as MTD)

When the effects of dose-limiting toxicity are less severe (non-fatal and treatable), some investigators have tried a more drastic version, the fast-track design, which starts with only one patient

Fast-Track Design

- This design was created to move through low doses using fewer patients. The design uses cohorts of one or three patients, escalates through the sequence of doses using a one-patient cohort until the first DLT is observed. After that, only three-patient cohorts are used.
- When a DLT is observed in a one-patient evaluation of a dose, the same dose is evaluated a second time with a cohort of three new patients, if no patient in this cohort experiences a DLT, the design moves to the next higher dose with a new cohort of three patients. From this point, it progresses as a standard design.

ABOUT FAST-TRACK DESIGN

The use of a fast-track design seems attractive because most clinicians want to proceed, as fast as they can. The fast-track design quickly escalates through early doses, thereby reducing the number of patients. On the other hand, the fast-track design might allow a higher percentage of patients to be treated at very high toxic doses; and it uses a single-patient cohort until the first DLT is observed seems too risky for some investigators. However, it might not be so – all would depend on the experience of the investigator trying it.

According to the (unstated) principle of "good medicine", each patient should be treated optimally: each patient should be treated with the "best" treatment that the doctor knows. According to this principle, each patient in phase I trial should be given a dose equal to the MTD - if the doctor knows what it is. In most cases, doctors may not know what is the MTD but they all "know" that, according to the standard design, the first few doses are likely "below" the MTD.

Despite all of those weaknesses, the standard design is still widely used in practice (70% to 80% of all phase I trials) because of its simplicity in logistics for the clinical teams to carry out (most of the times responsibilities fall on nurses, not doctors)

ST DESIGN: COMMON CRITIQUES

- Patients enter early are likely treated sub-optimally; may be we need to move up faster (against the principle of “good medicine”!)
- Only few patients left when MTD reached, not enough to estimate MTD’s toxicity rate (against the principle of “good statistics”!)

The standard design is not robust; expected rate of the selected MTD is strongly influenced by the doses used. If the trial is such that there are many dose levels below the MTD then the standard design will choose a dose far too low with greater probability than if there are fewer dose levels below the MTD.

Standard Design is SAFE, i.e. few patients are exposed to and died because of toxicities.

However, "safe" does not necessarily mean "good" – what's good for common, healthy people might not be good for patients; if not given enough medication, the patient would be killed by the cancer/disease.

According to the (unstated) principle of "good medicine", each patient should be treated optimally: each patient should be treated with the "best" treatment that the doctor. According to this principle, each patient in phase I trial should be given a dose equal to the MTD - if the doctor knows what it is. In most cases, doctors may not know what is the MTD but they all "know" that, according to the standard design, the first few doses are likely "below" the MTD.

According to “principle of “good medicine”, the patient should be treated with the best treatment the doctor knows. Patients enter early to a Phase I trial with Standard Design are likely treated sub-optimally; they receive a treatment level that the attending physician knows to be inferior. Some of these patients would likely die before any other therapy can be attempted. The newer design, the “continual reassessment method (CRM)” is an attempt to correct that by giving each patient a better chance of a favorable response.

In addition to the attempt to treat each patient more ethically, the CRM also updates the information of “the dose-response relationship” as observations on DLT become available and then to use this information to concentrate the next step of the trial around the dose that might correspond to the anticipated target toxicity level. It does so using a Bayesian framework, even though it has been argued that the CRM could be explained by likelihood approach.

The CRM is very attractive and has fostered a heated debate or debates which last for more than a decade. There are many variations of the CRM, we'll describe here a scheme based on a specific prior; the principle and the process are the same if another model is selected.

CRM is a Bayesian method

In most statistical inference problem, a parameter θ is considered to be a fixed but unknown constant. In a sub-area of statistics, parameters are considered as a random variable with a known probability distribution. This distribution is denoted, say, by $\pi(\theta)$ and called a **prior distribution**.

A probability function f(x;θ) could be represented as a conditional distribution with "variable" θ fixed:

$$f(x;\theta) = f(x \mid \theta)$$

The joint distribution of X and θ is re-formulated as

$$f(x,\theta) = \pi(\theta) f(x \mid \theta)$$
$$= g(x) h(\theta \mid x)$$

g(x) is the marginal density of X and h(θ|x) denotes the conditional density of θ, given the data X=x. This is called the posterior distribution of θ.

According to Bayesian Method, after data have been collected, a parameter θ is estimated by the mean of its posterior distribution.

CONTINUAL REASSESSMENT METHOD

Step 1: Choose the "maximum tolerated level" θ, the toxicity rate at the recommended dose level or MTD's (say, θ=.33 or whatever); this is a basic difference with standard design (SD).

Step 2: Choose a fixed number of patients to be enrolled; usually n = 19-24; this is another difference with SD (where the number of patients needed is variable).

Step 3: The CRM uses binary response (DLT or not); Let Y be the binary response such that Y=1 denote the occurrence of a pre-defined DLT. Let

$p(x) = Pr[Y=1|x]$ and

$\text{logit}\,[p(x)] = \log\{p(x)/[1-p(x)]\}$

The next step is to choose a statistical model representing the relationship between Y and dose level; for example, it could be described by the logistic (or probit) model :

$\text{logit}\,[p(x)] = \alpha + \beta x$

where x is the log of the dose d; or x is dose d.

Step 4: Use the baseline response/toxicity, adverse-effect rate (dose = 0) to calculate and fix the "intercept" α.

Step 5: Under the Bayesian framework, choose a prior distribution for the "slope" β; for example, "unit exponential" - one with probability density function $g(\beta) = \exp(-\beta)$

Step 6: From the model: logit [p(x); β] = α + βx, with β placed at "the prior mean" and set p(x) equal to the target rate θ, solve for dose x. This is dose for the first patient, a dose determined to reflect the current belief of the investigator/doctor as the dose level that produces the probability of DLT closest to the target rate θ - the "maximum dose with an acceptable and manageable safety" . This step fits the "principle of good medicine"! - the patient is treated at the MTD.

Step 7: After the first patient's toxicity/adverse-effect result becomes available, the "posterior distribution" of β is calculated and the posterior mean of β is substituted in logit $[p(x); \beta] = \alpha + \beta x$.

The next patient is treated at the dose level x whose probability p(x) is the target rate θ (with calculated posterior mean of β). This step is repeated in subsequent patients every time toxicity/adverse-effect result becomes available and the posterior distribution of β is re-calculated.

There are more than one ways to calculate the "posterior mean", one of which can proceed as follows – without going through the posterior distribution:

From the model "logit $[p(x_i; \beta)] = \alpha + \beta x_i$", the (Bernoulli) likelihood at dose i is:

$L(\beta) = p(x_i; \beta)^{\delta i} [1-p(x_i; \beta)]^{1-\delta i}$

And the mean β of is calculated from:

$$E(\beta) = \frac{\int \beta L(\beta) g(\beta) d\beta}{\int L(\beta) g(\beta) d\beta}$$

For quick and convenient use, we would need a software program for this calculation; a few versions of freeware are available.

Finally the MTD is estimated as the dose level for the hypothetical (n+1)th patient; n has been pre-determined, usually 19-24.

Use of CRM is on the rise; but it is not critique-free. The following are two typical concerns

First, the CRM might start the trial with an initial dose far above the "customary" lowest dose that is often one-tenth the LD10 in mice. This possibility makes many clinicians and regulatory agencies (e.g. FDA) reluctant to implement the CRM. After all, this is the first trial in human, little is known about the dose range - except results from animal studies. Some might go higher at the first dose, but not more than one-third the LD10 in mice. Some proposed that the trial always starts with the lowest dose as the dose for the first patient; CRM would start with the second dose/patient.

Secondly, there is a possibility that dose could be escalated for more than one dose level at a time (traditionally, as in standard design, doses are equally-spaced on the log scale or following a modified Fibonacci sequence with increases of 100, 67, 50, 40, and 33% for fifth and subsequent doses). Moller (1995) gave an example showing that the first dose could be escalated to the top level when the first patient has no LTD.

The strength of the CRM are still its three properties:

(1) it has a well-defined goal of estimating a percentile of the dose-toxicity relationship,
(2) it should converge to this percentile with increasing sample sizes, and
(3) the accrual is pre-determined. The standard design does not have these characteristics.

Suggested Exercises:

#1. When Phase I Cancer Trial following the Standard Design reaches a dose level with a toxicity rate 0f 40%, what is the probability that it would pass to the next higher dose?

#2. Consider a Phase I Cancer Trial with three doses (and toxicity rates): 15%, 35%, and 55%. Using the Standard Design, what is the probability that a subject would be treated at the last dose