Math 654: Design and Analysis of Clinical Trials

Chapter 1 Introduction to Clinical Trials

Wenge Guo

November 9, 2010

1. Introduction to Clinical Trials

CLINICAL TRIALS: The evaluation of intervention (treatment) on disease in a controlled experimental setting.

- The comparison of AZT versus no treatment on the length of survival in patients with AIDS
- Evaluating the effectiveness of a new anti-fungal medication on Athletes foot
- Evaluating hormonal therapy on the reduction of breast cancer (Womens Health Initiative)

Definition of a Clinical Trial (1)

- A clinical trial is a study in human subjects in which treatment (intervention) is initiated specifically for therapy evaluation.
- A prospective study comparing the effect and value of intervention against a control in human beings.
- A clinical trial is an experiment which involves patients and is designed to elucidate the most appropriate treatment of future patients.
- A clinical trial is an experiment testing medical treatments in human subjects.

Definition of a Clinical Trial (2)

- In **clinical trials**, the control group is the group of people who are on best current standard therapy or on no active intervention.
- The treatment may be prophylactic, diagnostic or therapeutic agents, devices, regimens, procedures, etc.
- At baseline, the control group must be sufficiently similar in relevant respects to the intervention group so that differences in outcome may reasonably be attributed to the action of the intervention.
- The experimental subjects are humans not animals, so ethics factor is very important and we must obtain informed consent from participants.

History of Clinical Trials

Historical perspective

- Historically, the quantum unit of clinical reasoning has been the case history and the primary focus of clinical inference has been the individual patient. Inference from the individual to the population was informal. The advent of formal experimental methods and statistical reasoning made this process rigorous.

- By statistical reasoning or inference we mean the use of results on a limited sample of patients to infer how treatment should be administered in the general population who will require treatment in the future.

- **Early History**

  1600 East India Company (A British company founded in 1600)

  In the first voyage of four ships– only one ship was provided with lemon juice. This was the only ship relatively free of scurvy.

- **1753 James Lind** (British doctor, Father of Nautical Medicine)

  "I took 12 patients in the scurvy aboard the Salisbury at sea. The cases were as similar as I could have them... they lay together in one place... and had one common diet to them all...

  To two of them was given a quart of cider a day, to two an elixir of vitriol, to two vinegar, to two oranges and lemons, to two a course of sea water, and to the remaining two the bigness of a nutmeg. The most sudden and visible good effects were perceived from the use of oranges and lemons, one of those who had taken them being at the end of six days fit for duty... and the other appointed nurse to the sick...

- **1794 Rush** (American doctor) *Treatment of yellow fever by bleeding*

  "I began by drawing a small quantity at a time. The appearance of the blood and its effects upon the system satisfied me of its safety and efficacy. Never before did I experience such sublime joy as I now felt in contemplating the success of my remedies... The reader will not wonder when I add a short extract from my notebook, dated 10th September. "Thank God", of the one hundred patients, whom I visited, or prescribed for, this day, I have lost none."

⋆ **Louis** (French physician): Lays a clear foundation for the use of the *numerical method* in assessing therapies.

Louis (1835) studied the value of bleeding as a treatment of pneumonia, erysipelas and throat inflammation and found no demonstrable difference in patients bled and not bled. This finding contradicted current clinical practice in France and instigated the eventual decline in bleeding as a standard treatment. Louis had an immense influence on clinical practice in France, Britain and America and can be considered the founding figure who established clinical trials and epidemiology on a scientific footing.

Table 2: *Pneumonia: Effects of Blood Letting*

Days bled after onset	Died	Lived	proportion surviving
1-3	12	12	50%
4-6	12	22	65%
7-9	3	16	84%

In 1827: 33,000,000 leeches were imported to Paris.

In 1837: 7,000 leeches were imported to Paris.

- **Modern clinical trials**:
  - ⋆ The **first** clinical trial with a properly randomized control group was set up to study streptomycin in the treatment of pulmonary tuberculosis, sponsored by the Medical Research Council, 1948 (UK). This was a multi-center clinical trial where patients were randomly allocated to streptomycin + bed rest versus bed rest alone.

    The evaluation of patient x-ray films was made independently by two radiologists and a clinician, each of whom did not know the others evaluations or which treatment the patient was given.

    Both patient survival and radiological improvement were significantly better on streptomycin.

⋆ **The field trial of the Salk Polio Vaccine**:

In 1954, 1.8 million first to third graders participated in the trial to assess the effectiveness of the Salk vaccine in preventing paralysis or death from poliomyelitis.

Incidence is low (1 in 2000).

Randomized component: 0.8 million children were randomized in a **double-blind placebo-controlled** trial.
Result: Incidence of polio in treated group is less than half of that in the control group.

Non-randomized component: Second graders were offered vaccine and first and third graders were formed control group.
Result: similar.

However, it turned out that the incidence of polio among children (second graders) offered vaccine and not taking it (non-compliers) was different than those in the control group (first and third graders).

Question: were treated children (second graders) and the control (first and third graders) similar?

- **Government sponsors clinical trials**: NIH (National Institutes of Health)
  - ⋆ NHLBI- (National Heart Lung and Blood Institute) funds individual and often very large studies in heart disease.
  - ⋆ NIAID- (National Institute of Allergic and Infectious Diseases) Much of their funding now goes to clinical trials research for patients with HIV and AIDS.
  - ⋆ NIDDK- (National Institute of Diabetes and Digestive and Kidney Diseases). Funds large scale clinical trials in diabetes research.

- **Pharmaceutical Industry**:
  - ⋆ Before World War II no formal requirements were made for conducting clinical trials before a drug could be freely marketed.
  - ⋆ In 1938, animal research was necessary to document toxicity, otherwise human data could be mostly anecdotal.
  - ⋆ In 1962, it was required that an "adequate and well controlled trial" be conducted.
  - ⋆ In 1969, it became mandatory that evidence from a randomized clinical trial was necessary to get marketing approval from the Food and Drug Administration (FDA).
  - ⋆ More recently there is effort in standardizing the process of drug approval worldwide. This has been through efforts of the International Conference on Harmonization (ICH).
    website:
    http://www.pharmweb.net/pwmirror/pw9/ifpma/ich1.html
  - ⋆ There are more clinical trials currently taking place than ever

before. The great majority of the clinical trial effort is supported by the Pharmaceutical Industry for the evaluation and marketing of new drug treatments.

2 Phase I and Phase II Clinical Trials

Phases of Clinical Trials:

- Preclinical (drug discovery): experimentation before a drug is given to human subjects
  - ⋆ lab testing for biologic activity (in vitro)
  - ⋆ testing on animals (in vivo)

- Clinical:
  - ⋆ **Phase I**: To explore possible toxic effects of drugs and determine a tolerated dose for further experimentation. Also explore pharmacology of the drug and investigate its interaction with other drugs, food and alcohol (these can be parallel to phase II-phase III trials). First-in-human trials.
  - ⋆ **Phase II**: Screening and feasibility by initial assessment for therapeutic effects; dose finding and further assessment of toxicities (safety and tolerability).
  - ⋆ **Phase III**: Comparison of new intervention (drug or therapy) to the current standard of treatment; both with respect to efficacy and toxicity.
  - ⋆ **Phase IV**: (post marketing) Observational study of morbidity/adverse effects.

2.1 Phase I clinical trials

Broad definition: Phase I trials are the first studies in which a new drug is administered to human subjects.

- Previous studies in the laboratory (in vitro)
- Previous studies in animals, e.g. rats, dogs (in vivo)

Objectives: Before efficacy (activity of a drug on disease) of the drug may be established, first must

- Determine a "safe," "tolerable" dose (through dose-escalation)
- Develop an appropriate schedule of administration
- Gain understanding of the *pharmacology* of the drug
- Need to examine interaction effects (drug-drug, drug-food, drug-alcohol) for safety profile and proper labeling.

In addition: Do this in a timely manner, using a small number of subjects.

Features:

- Most are not comparative but rather are "informational"
- "Interaction studies" are comparative, but not aimed for efficacy, still "informational" (for safety).

Types of studies:

- "Dose-finding" studies – determine the maximum dosage that can be given without serious "problems" – these studies are often themselves called "phase I studies" (especially for cancer studies)
- Clinical pharmacology studies – determine the $pharmacokinetics$ of the drug to aid in setting dosage schedules, understanding how "problems" are related to amount of drug present
- Drug-drug interaction studies – determine how other commonly used drugs affect important PK parameters, using a cross-over or parallel design
- Drug-food interaction studies – determine how food affects important PK parameters, using a cross-over or parallel design
- Alcohol/benzo interaction studies – determine how alcohol/benzo affects PD parameters, using a cross-over or parallel design

Pre-clinical Studies

Before administration to humans:

- *In vitro* studies – laboratory investigations using biological material but not actual organisms; look for biologic activity of the drug
- Dose-finding studies in rodents, large animal species (e.g. dogs)
- Pharmacology studies in rodents, dogs, etc
- Goal – "scale up" previous results to provide first idea of behavior in humans
- Advantage – Ethical issues involved in human experimentation circumvented (coming up)

Phase I Dose-finding Studies

Paradox: Although results of "dose-finding" will be carried forward to be used in studies of efficacy (phase II) and later to comparative trials (phase III) and eventually to routine patient care

- Standard approaches to design and analysis have little statistical justification
- Many texts on clinical trials devote little or no discussion to dose-finding (or pharmacology) studies, e.g. Freedman et al. mention only on p. 3–4!

Toxicity: "Problems" that may arise in direct response to administration of the drug – side effects

- Nature depends on the drug
- E.g. change in organ function – a drug to treat cancer may induce irregular heartbeat
- May be life-threatening and irreversible
- May be life-threatening and reversible
- May be non-life-threatening

Characterization: Often done on a standard, "graded" scale (especially for cancer research)

- Ordered categories increasing in severity – Grades I – IV or V
- Examples:

	I	II	III	IV
Abdominal Pain	Mild	Moderate (No trt)	Moderate (Trt)	Severe (Hospital)
Creatinine Clearance (cc/min/1.73 m^2)	60–75	50–59	35–49	< 35

Defining toxicity:

- Which toxicities are relevant depends on intervention, nature of likely subjects, clinical judgment – must be defined by the investigators
- What degree of toxicity is "acceptable" must be established

Terminology: *Dose Limiting Toxicity* (DLT)

- Serious or life-threatening but reversible
- Often used as the definition for dose-finding in cancer research

Assumption: Maximum benefit occurs at maximum doses

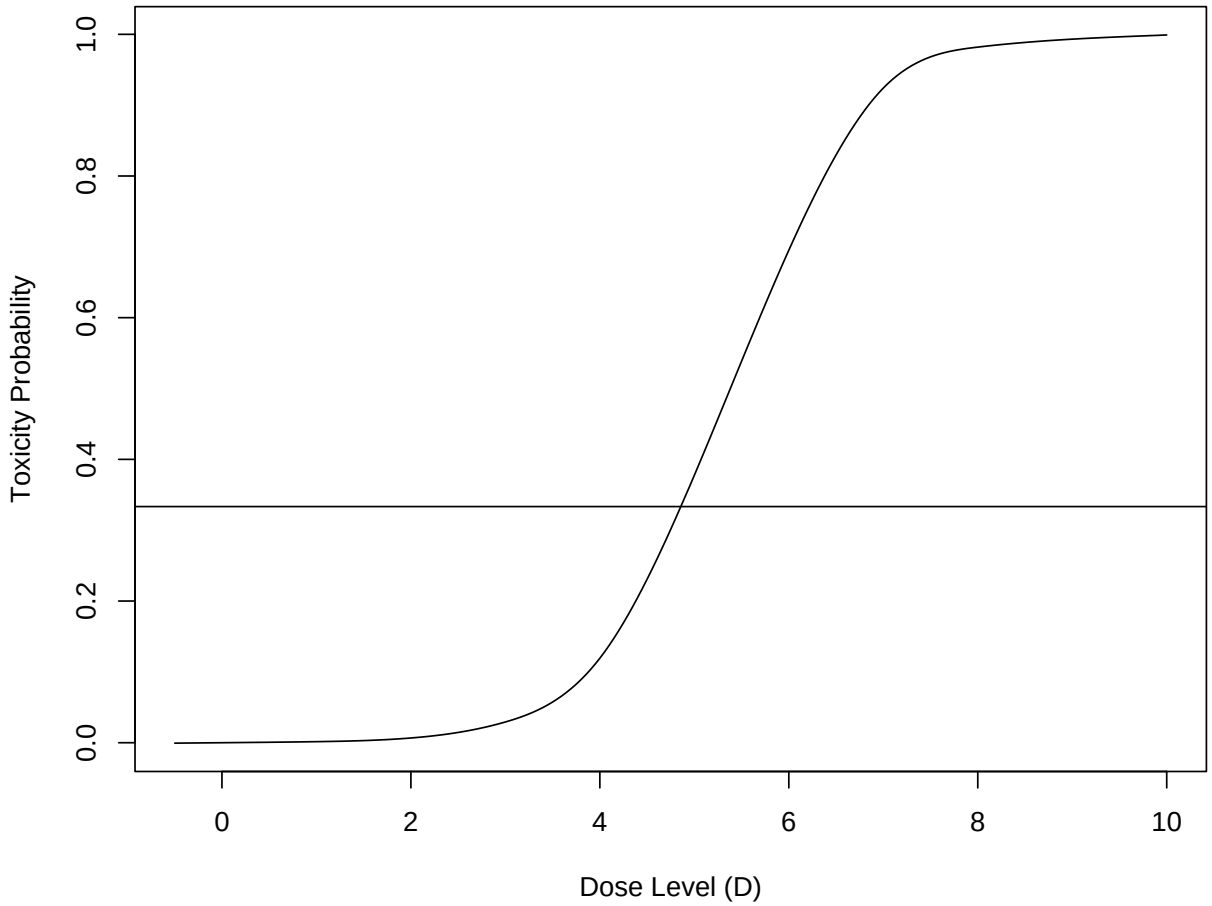
Main objective: Determine the *Maximum Tolerated Dose* (MTD)

- The "highest possible while still tolerable" dose
- Involves some level of toxicity we are willing to tolerate because of the drug's potential benefit
- E.g. "the dose that produces toxicity of grade III or worse in not more than one out of three patients" (often used for anticancer agents)

Statistically speaking: Determine the dose at which the proportion of subjects in the population who would develop toxicity if given this dose is 1/3

- I.e. the 33rd percentile of the distribution of toxicity in the population

Illustration of MTD: a hypothetical example



Subjects: Nature of subjects used depends on drug

- Healthy volunteers – used where toxicities unlikely to be severe, e.g. topical agents
- Patients with advanced disease – used where subjects have failed other therapies, toxicities likely. e.g. chemotherapy

Goal: Establish appropriate dose quickly while exposing as few subjects as possible to suboptimal doses that are likely not to be efficacious

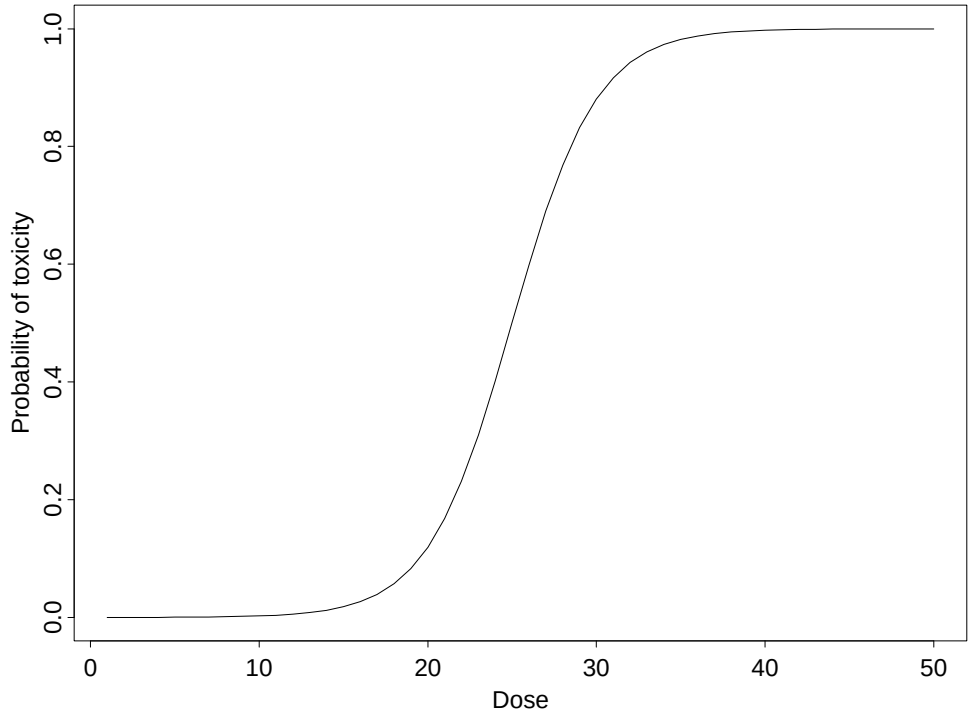
Ideal design:

- Select doses $D_1, \ldots, D_k$ for study such that one of these is (close to) the MTD
- Randomize subjects to each dose, n_i subjects at dose i
- Observe number r_i exhibiting DLT at each dose, and calculate proportions exhibiting DLT $p_i = r_i/n_i$
- Model dose-response (probability of toxicity) and fit to observed proportions at each dose
- Estimate MTD from fitted model

Models for dose-response:

- $Y = 1$ if toxicity observed for a subject at dose d
- $P(Y = 1|d) = f(d, \boldsymbol{\beta})$, f monotone in dose
- Dose d_0 associated with specified probability $p_0 = P(Y = 1|d_0)$ is $d_0 = f^{-1}(p_0, \boldsymbol{\beta})$
- E.g. If MTD is defined as dose where toxicity is 33%, $MTD = f^{-1}(0.33, \boldsymbol{\beta})$
- Estimate $\boldsymbol{\beta}$ from data and use to estimate MTD, i.e. $\widehat{MTD} = f^{-1}(0.33, \hat{\boldsymbol{\beta}})$
- More generally
  - ⋆ May estimate dose associated with any probability
  - ⋆ May estimate probability associated with any fixed dose
- Standard approach for animal experiments

$$P(Y=1|d)=\frac{\exp(\beta_0+\beta_1 d)}{1+\exp(\beta_0+\beta_1 d)},\ \ \boldsymbol{\beta}=(\beta_0,\beta_1)=(-10,0.4)$$



Problem: This approach not feasible in human subjects due to *ethical considerations*

- Because drug not previously used in humans, must test at lower doses first before feeling confident enough to move to higher doses
- Do not want to treat many subjects at dose that is too low to do good or too high (toxic)
- Result: cannot simply randomize subjects to different dose levels

Approach: "Adaptive" designs – *dose escalation*

- Try a dose in several subjects
- If no toxicities, try a higher dose in several (new) subjects
- Continue until dose is found that yields toxicity
- Sometimes, may "de-escalate" from a dose that is not tolerated
- Many variations on this idea

Result: Sample size is not specified in advance; rather, it is an outcome of the study

- Usually, sample size is small (~ 20)

Standard design (to find MTD): Select a sequence of increasing doses. Start at lowest dose level and administer drug to 3 subjects

1. If no toxicity observed in any of the 3 subjects, escalate to the next dose and begin again
2. If toxicity observed in 2 or more of the 3 subjects, STOP
3. If toxicity observed in exactly 1 subject of the 3, treat 3 additional subjects at this dose. If none of the additional subjects exhibits toxicity, then escalate to the next dose and begin again; otherwise, STOP

MTD: Usually defined as one of

- Dose at which trial stops
- Next lowest dose in the sequence
- Some fraction of the last dose

Determining the initial dose: The starting dose of the sequence is chosen in a conservative way, e.g.

- From rodent studies, estimate of LD_{10} is available; LD_{10} = dose where percentage of rodents exhibiting mortality is 10%
- Use as starting dose 1/10 of the rodent LD_{10} given on a mg/kg basis (scaled from rodent to human size)
- OR from larger animal (e.g. dog) studies, determine the $toxic\ dose\ low$ (TDL) = the lowest dose at which any toxicity seen
- Use as starting dose 1/3 of the dog TDL

Determining the sequence: Want to select doses in a way that will reveal the "MTD" without requiring an excessive number of dose levels

- Common technique – "modified Fibonacci" sequence
- Usual Fibonacci sequence 1, 1, 2, 3, 5, 8, 13, 21, ...
- Ratio of successive terms $\rightarrow 1.618$ ($\approx$ 62% increase)
- Modify to increase less rapidly with decreasing increments
- Alternative: equally-spaced doses on log scale over range

Example: Modified Fibonacci with $D =$ initial dose

Step	Usual	Modified	% Increment
1	D	D	–
2	$2 \times D$	$2 \times D$	100
3	$3 \times D$	$3.3 \times D$	67
4	$5 \times D$	$5 \times D$	50
5	$8 \times D$	$7 \times D$	40
6	$13 \times D$	$9 \times D$	29
7	$21 \times D$	$12 \times D$	33
8	$34 \times D$	$16 \times D$	33

Performance of the standard design: Example

Dose level	Actual toxicity prob. (π_i) (unknown in practice)	Prob (p_i) of stopping at this dose
1	0.15	0.186
2	0.20	0.237
3	0.25	0.231
4	0.30	0.178
5	0.33	0.096
6	0.50	

- E.g. probability of stopping at dose level 1 ($\pi_1 = 0.15$):

$$
\begin{aligned}
p_1 &= P[(Z_1 > 1) \cup (Z_1 = 1, Z_2 > 0)] \quad (Z_1, Z_2 \overset{iid}{\sim} \text{binomial}(3, \pi_1)) \\
&= P[Z_1 > 1] + P[Z_1 = 1, Z_2 > 0] \\
&= P[Z_1 > 1] + P[Z_1 = 1]P[Z_2 > 0] \\
&= 3\pi_1^2(1 - \pi_1) + \pi_1^3 + 3\pi_1(1 - \pi_1)^2(1 - (1 - \pi_1)^3) = 0.186
\end{aligned}
$$

- Given the trial reaches level 2, the probability of stopping at dose level 2 can be calculated in the same way (replace π_1 by π_2), which is 0.2912.
- The probability of stopping at dose level 2 ($\pi_2 = 0.20$):

$$p_2 = 0.2912 \times (1 - p_1) = 0.237.$$

- Chance of ever reaching the 33rd percentile is only 16.8% $(1 - p_1 - p_2 - p_3 - p_4)$
- Given the trial reaches the 33rd percentile (dose level 5), the chance of stopping there is only 57% (calculated similarly to p_1 by replacing π_1 with π_5)

Remarks: Statistically speaking

- If the true MTD is defined as previously, not clear that the dose announced as MTD is a credible estimate of this quantity
- The design has no intrinsic property that makes it stop at the 33rd or any other percentile of the toxicity distribution
- The announced MTD can only be at or near one of the doses in the sequence used, none of which may be exactly equal to the 33rd percentile
- No basis for accounting for sampling error (standard errors?)
- No appeal to formal statistical model
- Likely to treat most subjects at low doses

Remarks:

- The standard design with this method of declaring the MTD is widely used, despite statistical concerns
- The method of MTD determination is favored by investigators (the declared MTD depends on individual patient outcomes)
- The method of MTD determination is of concern to statisticians (sampling error is not taken into account)
- Proposals in the statistical literature for more rigorous analysis and other designs have been made

Formal approach to analysis of the standard design:

- Assume a statistical model for probability of toxicity at dose d

$$P(Y = 1|d) = f(d, \boldsymbol{\beta})$$

  $Y = 1$ if subject exhibits toxicity, 0 otherwise

- For the ith group of 3 subjects, let $Z_i = \#$ of subjects experiencing toxicity, corresponding $X_i =$ dose level
- It is possible to write out the likelihood for the data (Z_i, X_i), $i = 1, \ldots, n$, where n is the (random) sample size
- The MTD as formally defined could then be estimated by maximizing likelihood in $\boldsymbol{\beta}$ and solving for MTD
- Writing down the likelihood requires a recursive conditioning
- **Important:** Properties of this are not as simple because sampling was *not random*

Attempts to improve upon the standard design and analysis:

Two-stage designs (Storer, 1989):

- Stage 1 – use very few patients to get to the dose region where the "action is" (close to MTD)
- Stage 2 – use a version of the standard design and find the MTD
- Allow dose de-escalation (go to lower dose) as well as escalation ("up and down" design)
- Use a statistical model, MLE to estimate MTD formally

Two-stage designs (Storer, 1989):

- Stage 1:
  - ⋆ Single subject at a dose
  - ⋆ If no toxicity, escalate dose for next subject
  - ⋆ If toxicity, de-escalate dose for next subject
  - ⋆ Begin second stage at first toxicity
- Stage 2:
  - ⋆ 3 subjects at a dose
  - ⋆ If no toxicity escalate dose for next group
  - ⋆ If 1 toxicity, add 3 subjects at this dose
  - ⋆ If > 1 toxicity, de-escalate dose for next group
  - ⋆ # groups fixed in advance

Stochastic approximation methods (Anbar, 1984):

- *Estimate* the next dose sequentially

Continual reassessment method (CRM, O'Quigley, Pepe, and Fisher, 1990):

- Bayesian approach – specify a prior distribution for the MTD and model for probability of toxicity
- First subject gets dose = prior value of the MTD
- After each subject, use Bayes' rule to update the posterior distribution of the MTD given the data so far
- Use the mode of this distribution (Bayesian "estimate") as dose for next subject
- Stop after a fixed number of subjects and do a Bayesian analysis to estimate the MTD

Disadvantages: Both approaches

- Require analysis after each subject
- May be too aggressive
- Require doses that may be difficult to prepare
- Modifications have been suggested; are area of current research and investigation