



# CLINICAL TRIALS: Part 2 of 2

---

Lance K. Heilbrun, Ph.D., M.P.H.  
Professor of Medicine and Oncology  
Division of Hematology and Oncology  
Wayne State University School of Medicine

Assistant Director, Biostatistics Core  
Karmanos Cancer Institute


[heilbrun@karmanos.org](mailto:heilbrun@karmanos.org)



## Definition of a clinical trial (CT)

---

- Prospective controlled experiment
- Human subjects with a specific disease or medical condition
- Asks an important research question
- Clinical event(s) as outcome(s)
- Done in clinic or medical setting
- Evaluates one or more interventions



# Phases (Types) of Clinical Trials

---

- |             |                              |                                |
|-------------|------------------------------|--------------------------------|
| <b>I:</b>   | <b>dose finding</b>          | <b>(Rx threshold)</b>          |
| <b>II:</b>  | <b>efficacy @ fixed dose</b> | <b>(safety &amp; efficacy)</b> |
| <b>III:</b> | <b>comparing treatments</b>  | <b>(randomized CT)</b>         |
| <b>IV:</b>  | <b>late/uncommon effects</b> | <b>(expanded safety)</b>       |



## Features of a typical Phase I CT

---

- Special type of pilot study (typically, N=20-30)
- Dose-finding for cytotoxic drugs (or XRT ?)
- After pre-clinical studies (in-vitro, in-vivo)
- First time new Rx used in humans
- For patients who have failed standard Rx's
- Usually done at a single institution



# Overall strategy for Phase I CT

---

- Low starting dose (very likely to be safe)
  - 1/10 of lethal dose to 10% (LD10) of mice
  - 1/10 LD10 in most drug-sensitive species
- Dose increased as new patients enter trial
- Side effects (toxicities) carefully monitored
- Seeking highest dose with reversible toxicities
- Graded via NCI Common Toxicity Criteria



## Goals of a Phase I CT

---

- Find maximum tolerated dose (MTD)
- Determine toxicity profile, safety limits
- Discover the dose-limiting toxicity (DLT)
- Characterize any anti-tumor activity
- Investigate clinical pharmacology of Rx
- Recommend dose level for Phase II trial



# Choosing dose levels (for adults)

---

## Pre-defined dose levels

- A) traditional “cohorts of 3” design
- B) accelerated titration (AT) design
- Doses determined by statistical modeling
  - modified continual reassessment method (mCRM design)



# Traditional “cohorts of 3” design

---

- Numerical escalation algorithm defines successive doses
- Doses are decreasing multiples of the previous dose
- Multiples derive from the Fibonacci sequence:  
1, 2, 3, 5, 8, 13, 21, ....
- Modified Fibonacci dose multiples become:  
2.00, 1.67, 1.50, 1.40, 1.33, 1.33, ....



# Traditional “cohorts of 3” design

---

- Enter 3 patients at starting dose
- Count # patients with a DLT:
  - if 0, enter 3 patients at next dose
  - if 1, enter 3 patients at current dose
  - if  $\geq 2$ , conclude MTD = previous dose



# Traditional “cohorts of 3” design

---

- Dose escalation stops when  $\geq 1/3$  of patients have DLT at a given dose level
- MTD is the next lower dose level
- Dose cohorts  $\geq 4$  sometimes used
- May require many dose levels to find MTD



# Accelerated titration (AT) design

---

- Several variations; all use std starting dose
- Start with 1-patient cohorts
- 40% or 100% dose ↑'s until first course DLT
- Then 3-patient cohorts with ↓ dose ↑'s
- Intra-patient dose ↑'s allowed
- Requires fewer patients (often, N = 15-20)

Simon et al. JNCI, 89:1138-47, 1997.



# mCRM design

---

- modified Continual Reassessment Method
- No pre-specified dose levels
- Starting dose calculated from PI's estimates of DLT probability at an assumed dose
- Dose-toxicity statistical model used to *calculate* next target dose level
- Usually, 3-patient dose cohorts



## mCRM design

---

- Model uses ALL accumulated dose/toxicity data before refitting model
- Identifies dose with “acceptable” probability (e.g., 0.20 – 0.40) of DLT

Piantadosi et al. Cancer Chemotherapy & Pharmacology, 41:429-436, 1998.



# Phase I CT of a biologic agent

---

- $\uparrow$  dose may not  $\Rightarrow$   $\uparrow$  biologic effect
- $\uparrow$  dose may not  $\Rightarrow$   $\uparrow$  toxicity
- Biologics may have low toxicity potential
- Seek the biologically effective dose (BED)
- Even w/o toxicity, doses  $>$  BED may  $\downarrow$  benefit
- Use fewer doses, but farther apart ?



## Randomization in a Phase I trial ?

---

- Not often used (safety issue)
- Unethical with cytotoxic agents
- Possibly useful for biologics with potential for low toxicity and high benefit
- Vaccine dose ↑ trial in renal cell CA found immune responses, with no DLT's:

Simons et al. *Cancer Res.*, 57:1537-46, 1997.



## Features of a typical Phase II CT

---

- Special type of pilot study (often, N=20-50)
- Get stat'l estimates of safety & efficacy at fixed dose found in Phase I trial
- For patients who have failed first choice Rx's
- Either single institution or multi-center
- Early termination option often allowed



# Designs for Phase II trials

---

- Fixed sample size
- Classical fully sequential:
  - monitor Rx effect after EACH patient
  - requires quick outcome, slow accrual
- Group sequential:
  - monitor Rx effect in stages of accrual
  - for slower outcomes or faster accrual



## Fixed sample size in Phase II CT

---

- Oldest (and simplest) design
- $N$  determined statistically in advance
- No formal opportunity for early stopping
- Requires larger  $N$  (on average), but that yields more statistical information
- Still used, but less common today



# Fully sequential designs

---

- Monitor Rx effect after EACH patient
- More cumbersome, labor intensive
- Difficult for multi-center implementation
- Rarely used for chronic diseases
- Not suitable for delayed endpoints



# Group sequential designs

---

- Monitor Rx effect in “stages” of accrual
  - 2 stages most often used
  - Determine whether TRUE success rate ( $p$ ) is:
    - “high enough” (e.g.,  $p \geq 0.40$ ) OR is
    - “too low” (e.g.,  $p \leq 0.20$ ) for further study
  - Want to distinguish regions where  $p$  is located
- Simon R. Controlled CT, 10:1-10, 1989.



## To distinguish $p \leq 0.20$ from $p \geq 0.40$

---

- Stage 1: enroll 17 patients; if  $\leq 3$  (18%) successes, stop trial (Rx not worthy).  
If  $\geq 4$  successes, begin Stage 2.
- Stage 2: enroll 20 more patients; if (among 37),  $\leq 10$  (27%) successes, stop trial.
  - If  $\geq 11$  (30%) successes, conclude  $p \geq 0.40$  and that Rx is worthy of further study.



## 2-stage Phase II CT design

---

- “Optimal”: minimizes mean N required
- Only stop early for “poor” Rx results
- Do not confuse sample estimate of success rate ( $s/n$ ) with TRUE (and unknown)  $p$
- $s/n$  is “evidence” on which to conclude that  $p \leq 0.20$ , or to conclude that  $p \geq 0.40$
- Can make 1 of 2 types of inferential errors



# Motivation for 2-stage Simon designs

---

- To halt use of an unworthy Rx ASAP, but based on convincing evidence
- Temporary CT closure after Stage 1 may be needed to evaluate recent patients
- k-stage designs ( $k \geq 3$ ) can be lengthy CT's
- Uncommon (e.g., in cancer) to stop early for Rx results that are "very good"



## Other 2-stage Phase II CT designs

---

- 2-dimensional early stopping rule
- Stop trial early if success rate is “too low”,  
OR if toxicity rate is “too high”
- More complex to explain, design, implement
- Elegant, but not widely used yet

Bryant & Day. Biometrics, 51:1372-83, 1995.



## Randomization in Phase II trials ?

---

- Phase II trial results of same Rx will vary
- Rand. better ensures comparable patients if several Phase II trials open at once
- Direct comparison of Rx results NOT intended
- Goal is to select, with high probability, the Rx which is superior by amount "D"

Simon et al. *Cancer Treat. Reports*, 69:1375-81, 1995.



# Special types of Phase III trials

---

- Crossover studies
- Equivalence trials
- Cluster randomized trials



# Crossover studies

---

- Two Rx's (A, B) and two Rx periods (1<sup>st</sup>, 2<sup>nd</sup>)
- Two Rx sequences: A, B and B, A
- Patient is randomized to a Rx *sequence*
- All patients receive both Rx's
- Each patient is his/her own control
- Eliminates between-patient variability



# Crossover studies - advantages

---

- Require ↓ N than parallel groups RCT
- Both Rx's handled equitably via the Rx sequence randomization
- Masking can be used
- Good for chronic diseases having dramatic Rx results: asthma, migraine, rheumatism



# Crossover studies - disadvantages

---

- Carry-over effect (need “wash out” period)
- Drop-outs (do not get Rx # 2)
- Not suitable for diseases without dramatic, rapid Rx results: cancer, stroke, paralysis
- Less convenient for patient: 2 Rx's, more time

Senn S. Statistical Methods in Medical Research, 3:303-24, 1994.



# Equivalence trials (ET)

---

- RCT test whether Rx outcomes differ
- ET test whether Rx outcomes are “the same”
- “Same” means “close enough”, i.e., within an equivalence bound,  $d$
- RCT: do Rx results differ by  $\geq “d”$  ?
- ET: do Rx results differ by  $< “d”$  ?



# Equivalence trials (ET)

---

- Failure to detect a Rx difference in a RCT does NOT  $\Rightarrow$  equivalence of Rx effects
- “Absence of evidence is not evidence of absence.”
- Equiv. bound “d” usually  $<$  “ $\Delta$ ” used in RCT
- Directional hypothesis  $\Rightarrow$  “non-inferiority trial”
  - new Rx is at least as good as old Rx,
  - new Rx not worse than Old by  $>$  “d”



# Equivalence trials (ET)

---

- “Non-inferior” Rx’s should have other benefits:  
↓ toxicity, or ↓ cost, or ↑ convenience
- ET’s beneficial when Rx advances are difficult to achieve for a given disease
- ET may require ↑ N as compared to RCT
- Bioequivalence: a special type of ET, e.g., to show drug blood level is “equivalent” to std Rx. New Rx is then both safe & efficacious.



# Cluster randomized trials

---

- Some interventions not amenable to randomizing individual subjects
- Suppose: RCT to compare 2 smoking prevention programs for teenagers
- Cannot randomize individual students at a given school and “isolate” their exposure to only 1 health education msg
- “Contamination” of 1 Rx group by the other



# Cluster randomized trials

---

- Solution: randomize entire groups (“clusters”) of subjects; apply only 1 Rx to all members
- Example: 20 high schools (each with 1,000 students) are randomized to 2 interventions
- Effective sample size to compare 2 smoking prevention programs is now 10 vs 10
- To ↑ stat'l power, ↑ # and/or size of clusters



# Interpreting reports of CT's

---

- Does CT address important question ?
- Is study design appropriate ?
- Are stat'l endpoints relevant to question(s) ?
- Are stat'l endpoints well defined ?
- Are ALL patients accounted for ?
- Do results fit with clinical experience ?
- Should results change clinical Rx practice ?



# Summary of CT's discussed

---

- Phase I (dose-finding & toxicity; MTD, DLT)
- Phase II (safety & efficacy @ a fixed dose)
  - estimation of success & toxicity rates
- Special types of Phase III RCT's:
  - crossover, equivalence, cluster
- ALL CT's require IRB-approved *protocol*



## CT References – Phase I, II

---

- [www.clinicaltrials.gov/ct/info/resources](http://www.clinicaltrials.gov/ct/info/resources)
- Green, Benedetti, Crowley. Clinical Trials in Oncology, 2<sup>nd</sup> Ed. Chapman & Hall, 2003.
- Crowley (Editor). Handbook of Statistics in Clinical Oncology. Marcel Dekker, 2001.
- Piantadosi. Clinical Trials: A Methodologic Perspective. Wiley & Sons, 1997.



## References – special Phase III CT's

---

- Matthews. An Introduction to Randomized Controlled Clinical Trials. Oxford U. Press, 2000.
- Pocock. Clinical Trials: A Practical Approach. Wiley & Sons, 1983.
- Jones et al. Trials to assess equivalence: the importance of rigorous methods (with editorial). Brit Med J, 313:36-39, 1996.



# A final thought

---

Virtually all roads in medical research must eventually lead to a **clinical trial(s)** IF we are to learn whether those “roads” can improve human health.

Thank you.

Questions ?