Randomized Clinical Trials

Hematology/Oncology Lecture Series

Elizabeth G. Hill, PhD

Associate Professor of Biostatistics

17 November 2011

Outline

- Defining the clinical trial

- Ethics in human research

- Stages of trial design in drug development

- Statistical considerations in trial design

- Interim analyses

Clinical trial definition

A *clinical trial* is an *experiment* testing a medical treatment on human subjects.

Not all human research studies are *experimental*, and it is this characteristic that distinguishes the clinical trial from other forms of medical research.

But what do we mean by an experiment?

"...the essential characteristic that distinguishes experimental from nonexperimental studies is whether or not the scientist controls or manipulates the treatment (factors) under investigation." (Piantadosi, *Clinical Trials: A Methodologic Perspective*, 2005)

Common characterstics of clinical trials

Frequently investigators conducting experiments will make an effort to exert additional control over extraneous factors that may contribute to outcome variability (random error) or bias (systematic error). Therefore, the design of a clinical trial may implement:

- An internal control group
- Stratification based on known prognostic factors
- Randomization
- Explicit comparison

However, none of these features is required of a clinical trial.

Equipoise

"Equipoise is the concept that a clinical trial (especially a randomized trial) is motivated by *collective uncertainty about the superiority of one treatment versus its alternative.* The existence of equipoise helps to satisfy the requirement that study participants not be disadvantaged. It ... supports a comparative trial as the optimal course of action to resolve scientific uncertainty ..." (Piantadosi, 2005)

Equipoise (cont.)

And from the original ...

> "... at the start of the trial, there must be a state of clinical equipoise regarding the merits of regimens to be tested, and the trial must be designed in such a way as to make it reasonable to expect that, if it is successfully conducted, clinical equipoise will be disturbed." (Freedman, 1987)

The Nuremberg Code

Adopted in 1947 in response to attrocities committed by physicians in concentration camps of Nazi Germany

1. Subject participation is voluntary

2. There must be no reasonable alternative to conducting experiment

3. Anticipated results founded on biologic principles and animal (pre-clinical) experimentation

4. Unnecessary suffering and injury are avoided

5. No expectation of death or disability as a result of the study

6. Risk to patient is consistent with humanitarian importance of the study

The Nuremberg Code (cont.)

7. Subjects protected against possibility of death or injury

8. Conducted by qualified scientists

9. Patient can stop participation at will

10. Experiment should be terminated if injury seems likely

Helsinki Declaration

- Adopted in 1964 by World Medical Association as a formal code of ethics for physicians engaged in clinical research with revisions in 1975, 1983, 1989, 1996, 2000

- 2000 revision was in response to controversial AIDS trials in Africa conducted by US investigators

- Placebo-controlled trials investigating prevention of mother-to-child HIV transmission

- Ethical issue surrounding placebo control when drug therapy known to be effective in reducing transmission

- At the time, standard of care in Africa was 'no treatment'

Helsinki Declaration (cont.)

2000 revision:

> *"The benefits, risks, burdens and effectiveness of the method should be tested against those of the current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists."*

Belmont Report

Report in 1978 by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research outlining ethical principles and guidelines for protection of human subjects and articulated principles and actions of IRBs.

- Risks to study participants are minimized

- Risks are reasonable in relation to anticipated benefits

- Selection of study participants is equitable

- Informed consent is obtained and documented

- Adequate provisions for monitoring data collected to ensure subject safety

- Privacy and confidentiality

A contemporary perspective

Emmanuel et al.[†] synthesized traditional codes, declarations, and relevant literature on ethical conduct of human subject research. Added one more component in 2003.

1. Collaborative partnership

2. Social or scientific value

3. Scientific validity

4. Fair subject selection

5. Favorable risk-benefit ratio

6. Independent review

7. Informed consent

8. Respect for potential and enrolled subjects

[†]Emmanuel, Wendler and Grady (2000). What makes clinical research ethical? *JAMA* 283:2701-2711.

Stages of trial design in drug development

- Phase I
  - Dose-finding
  - Designed to find the highest safe dose
  - ≤ 30 patients

- Phase II
  - Safety and efficacy
  - Generally not a 'head-to-head' comparison
  - 20 - 80 patients

- Phase III
  - Definitive comparative trial against standard of care
  - Hundreds to thousands of patients

Phase I trial goals

- Classic Phase I trials
  - Find the highest dose that is deemed safe = maximum tolerated dose (MTD)
  - Monitor dose-limiting toxicities (DLTs)
  - Find highest dose that has a DLT rate of x% or less (in cancer usually 20% - 40%)
- Newer Phase I trials
  - Find the dose that is considered to be safe and has optimal biologic/immunologic effect (OBD)
  - Goal is to optimize "biomarker" response within safety constraints

Classic phase I assumptions

Efficacy and toxicity both increase with dose
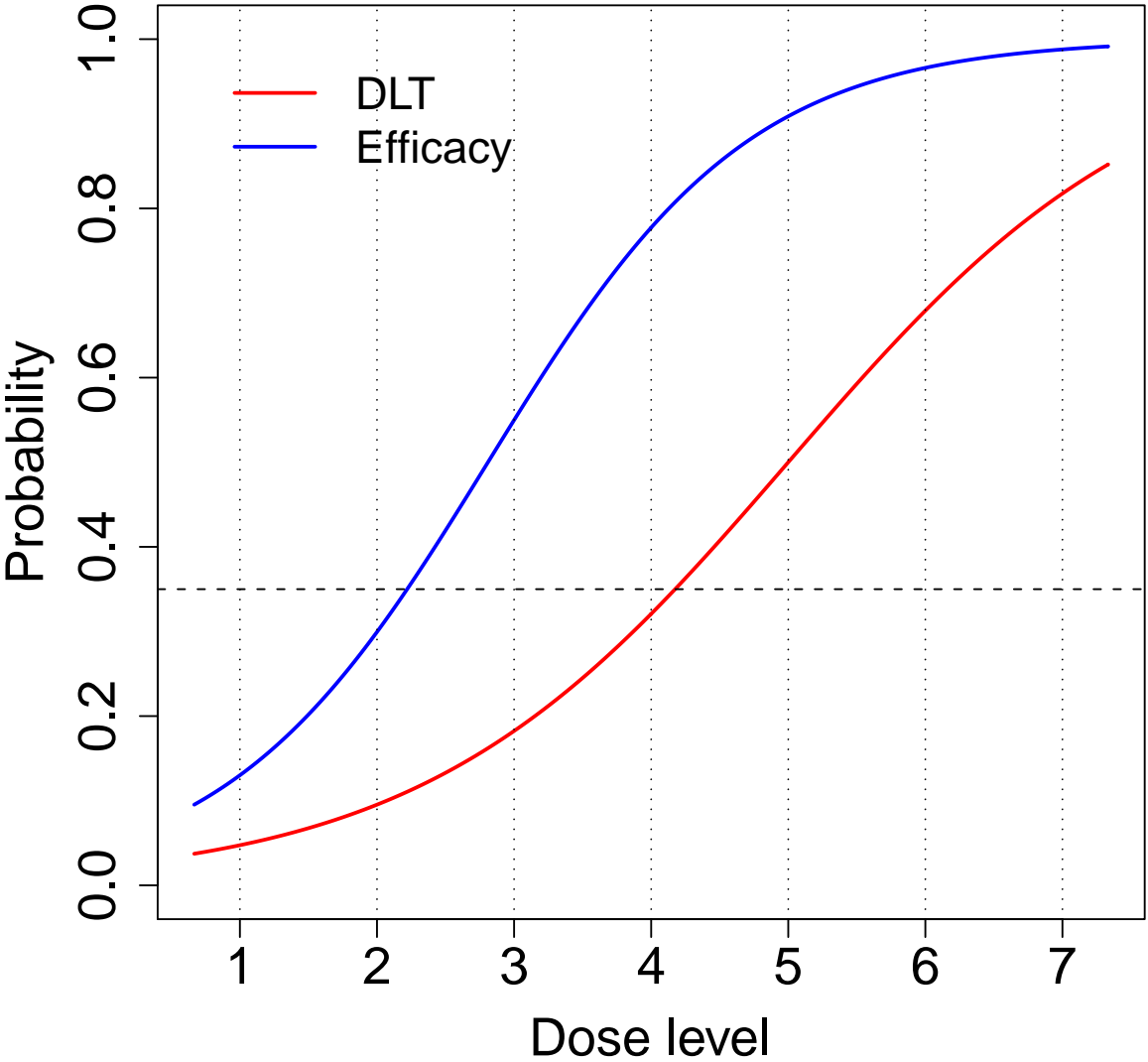
Classic phase I "3+3" design

- Pre-specify a set of (usually 3 - 10) doses to consider

> Treat 3 patients at dose k
>
> 1. If 0 patients experience a DLT, escalate to dose $k + 1$
>
> 2. If 2 or more patients experience a DLT, de-escalate to level $k - 1$
>
> 3. If 1 patient experiences a DLT, treat 3 more patients at level k
>
>   (a) If 1 of 6 experiences a DLT, escalate to dose $k + 1$
>
>   (b) If 2 or more of 6 experience a DLT, de-escalate to level $k - 1$

- MTD is considered the highest dose at which 1 or 0 out of 6 patients experiences a DLT

New paradigm in cancer - "targeted" therapy

- Agent selective for a molecular 'target'

- Disrupt carcinogenesis by interfering with specific pathway

- Considered less toxic than traditional cyotoxic agents

- Implications for phase I design
  - Toxicity may be extremely low
  - Need to ensure agent 'hits' the target
  - Efficacy may not increase monotonically with dose
  - MTD being replaced by OBD

Possible relationships for targeted therapies
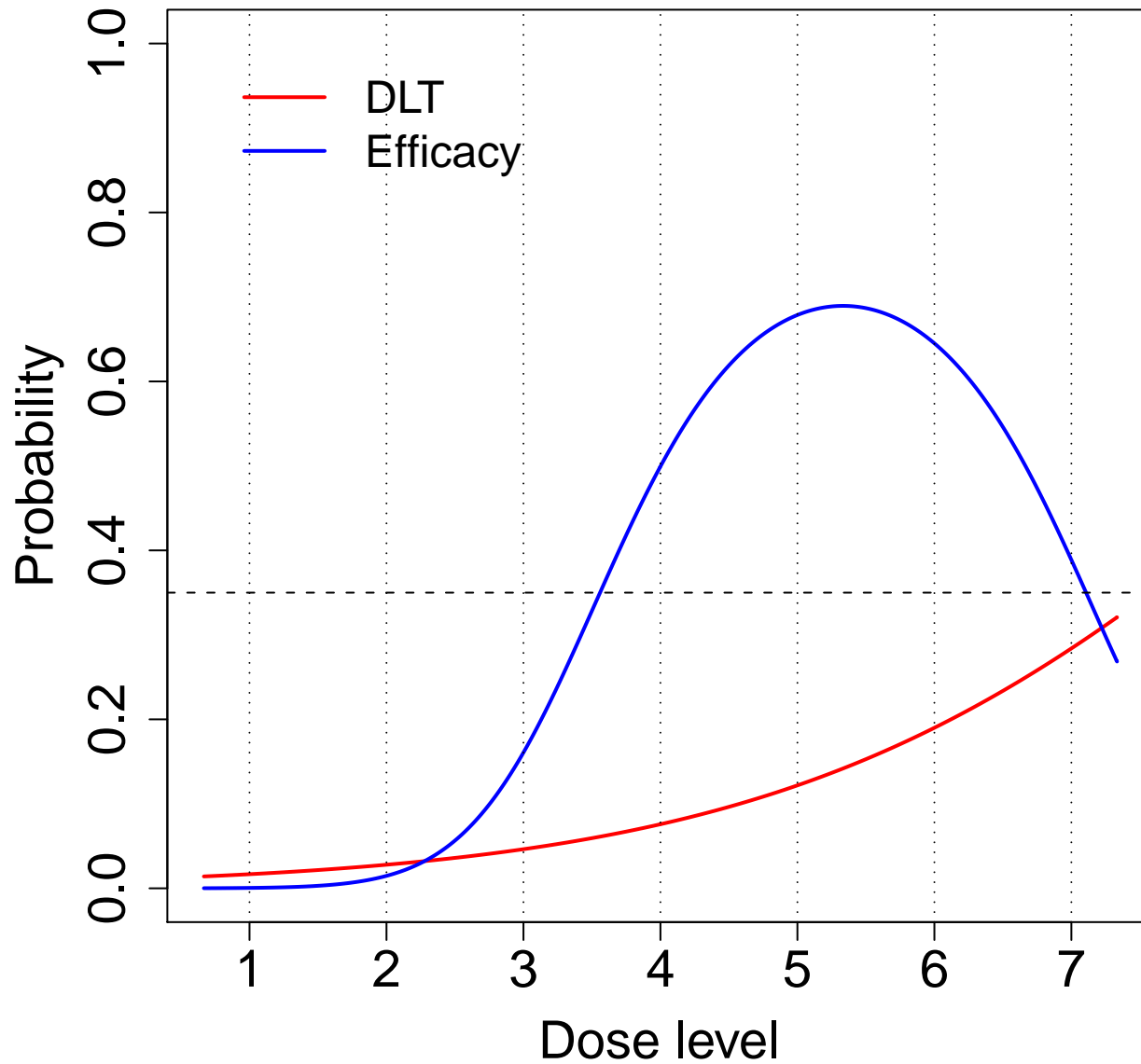
Phase II Safety and Efficacy trials

- Provide preliminary information on whether a treatment is efficacious

- Often single arm

- Small - only large effects are detectable

- Quick
  - Short-term endpoints preferred
  - Often the endpoint is surrogate for desired endpoint
  - E.g. Progression free survival or response rather than overall survival

- Often unblinded

- Sometimes randomized but rarely for head-to-head comparison

- Typically set $\alpha = 0.05 - 0.10$ and power = 80 - 90%

Phase II designs

- Single arm
  - Results compared to historical control rate
  - May be unsatisfying if control rate not well-defined
  - Study population may lack comparability to historical control population

- Randomized selection design
  - "Pick the winner"
  - Patients randomized to two (or more) arms
  - No head-to-head comparison of arms, but rather comparisons to null historical rate
  - Goal is to identify best dose/schedule/regimen to take forward into phase III when there is no *a priori* information that one is preferable

Phase II designs (cont.)

- Randomized with control arm
  - Control arm is included to ensure historical rate is "on target"
  - Control arm is *not* included for the purposes of head-to-head comparison (due to small sample size)

Phase III trials

- Comparative trial with two or more arms

- Goal is to show definitive clinical efficacy relative to current standard

- Conducted by large pharmaceutical companies or cooperative groups

- Large, expensive, long duration

- Usually multi-center
  - Infrastructure
  - IRB approval at each site

- Coordination with FDA

- Establish DSMB for oversight

- Typically set $\alpha \leq 0.05$ and power $\geq 80\%$

Phase III statistical design considerations

From page 2 of the Ohtsu article:

> "AVAGAST was a prospective, *random-assignment*, *double-blind*, *placebo-controlled* phase III clinical trial. The protocol was approved at each participating site by an independent ethics committee or institutional review board ... The trial was carried out in accordance with the Declaration of Helsinki. All patients provided written informed consent before study entry.
>
> Patients were assigned (*1:1 ratio*) to treatment by using *permuted-block randomization* ..., with geographic region (Asia-Pacific/Europe/Pan-America), fluoropyrimidine (capecitabine/FU), and disease status (metastatic/locally advanced) as *stratification factors*."

Bias and error

Control of *bias* and reduction of *random error* are two major objectives in statistical design considerations
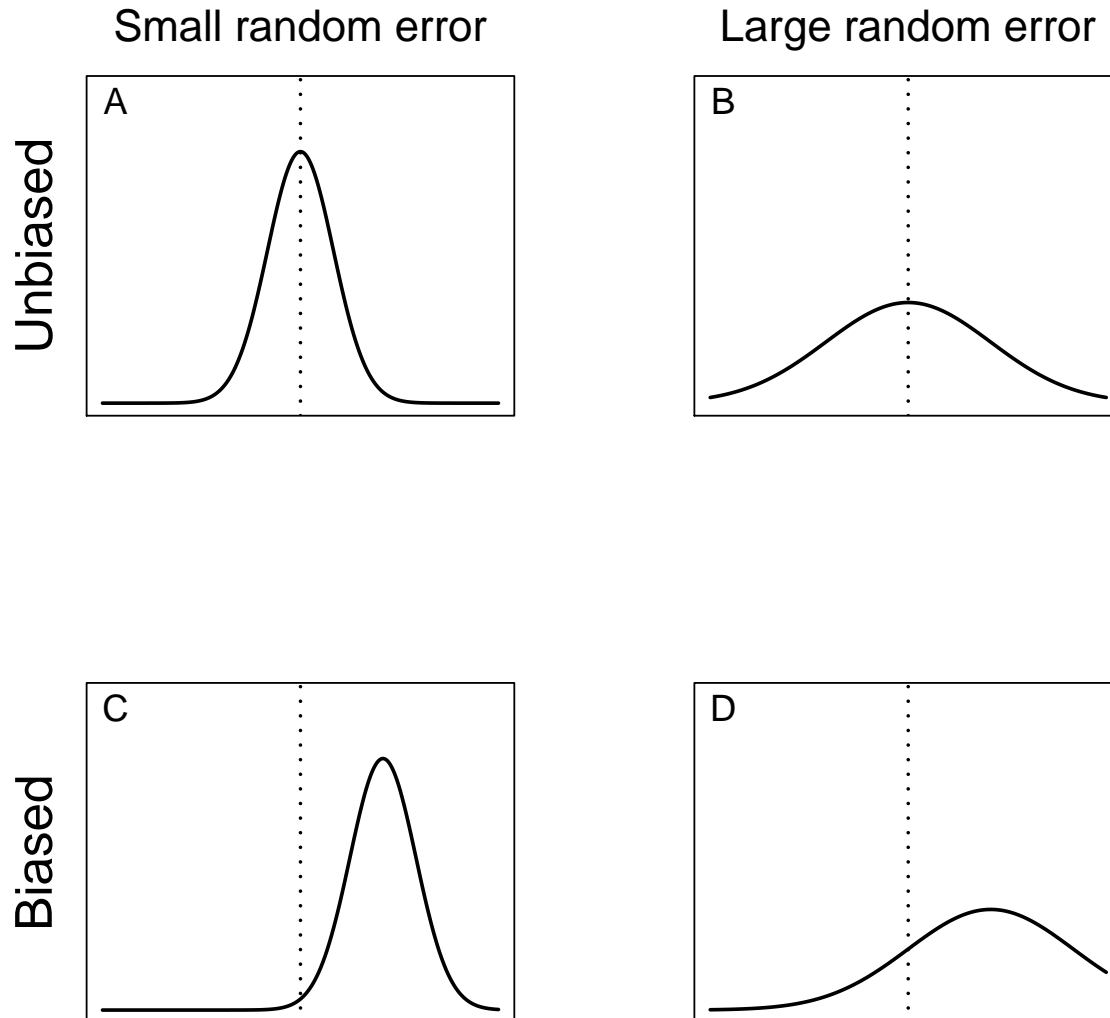
Sources of bias

1. *Selection bias*

   - Patients in arms systematically different with respect to prognostic factors
     - Bias the observed treatment effect
     - Can influence internal validity
   - Study cohort not representative
     - Can influence external validity
     - Compromises generalizability

2. *Treatment/procedure selection bias*

   - Healthier patients selected for a particular treatment
   - Systematic difference in composition of treatment groups
   - Can bias treatment difference

Sources of bias (cont.)

3. Postentry exclusion bias

- Inappropriate exclusion of eligible and enrolled subjects from the analysis

- Exclusion often due to seemingly reasonable clinical reasons

- Breaks the 'experimental paradigm'

- Example - subjects that fail to complete therapy are excluded from the analysis

- Example - subjects that die due to 'other' causes are excluded from an analysis of overall survival

Sources of bias (cont.)

4. Selective loss of data

- Loss of data resulting from unworkable or suboptimal outcomes or errors in study conduct

- Endpoint poorly selected for patient population

- Frequency of follow-up inappropriate for assessment of desired endpoint or is not followed as specified in protocol

- Example - patient population is seriously ill cohort and endpoint is based on patient self-report; endpoint may suffer from survivor bias

- Example - endpoint is time to progression; follow-up with patients every 6 months may be inadequate for accurate assessment

Sources of bias (cont.)

5. Assessment bias

- Patient self-assessment lacks objectivity

- Clinician assessment can be influenced by expectation of treatment effect

- Can bias endpoint in direction of prior expectation

6. Uncontrolled confounders

- Confounder is a variable that masks the true treatment effect

- Common confounders are age, race, gender, disease severity, comorbidities

- Example - treatment arm is significantly younger than placebo arm, and outcomes in older patients are more severe

Controlling for bias by design - Randomization

- Patients randomly assigned to treatment
- Controls for:
  - Selection bias
  - Treatment bias
  - Uncontrolled confounders
- Types of randomization
  - Simple
  - Permuted block
  - Stratified permuted block
  - Adaptive
  - Group
- Allocation ratio (1:1, 2:1, etc.)

Simple randomization

- Each new treatment assignment made without regard to assignments already made

Subject	Treatment
1	B
2	B
3	A
⋮	⋮
100	B

- On average, equal numbers of subjects will be assigned to each arm

- However, for any given trial, simple randomization can cause imbalance in number of subjects assigned to one of the treatments

Permuted block randomization

- Simple constraint to improve balance in treatment assignments

- Each "block" contains a pre-specified number of treatment assignments

- Block size must be an integer multiple of the number of treatments

- Two treatment groups (A and B) and blocks of size 4

Block	Treatment			
1	A	B	A	B
2	A	B	B	A
3	B	A	B	A
⋮	⋮			
25	B	B	A	A

Stratified permuted block randomization

- Strata defined by combinations of important prognostic factors

- In Ohtsu article (see slide 23), strata are defined by
  - Geographic region (Asia-Pacific, Europe, Pan-America) - Introduction details differences in disease stage by geography
  - Fluoropyrimiidine (capecitabine or FU) - Patients received chemotherapy + bevacizumab or chemotherapy + placebo, where chemotherapy = capecitabine-cisplatin or fluorouracil-cisplatin
  - Disease status (metastatic or locally advanced) - Differences in disease status would affect patient survival outcomes

Stratified permuted block (cont.)

- Permuted block randomization occurs within strata

- In the Ohtsu article there are 12 strata - Example: Asia-Pacific, FU, metastatic

- Ensures balance of prognostic factors across treatment arms

Adaptive randomization

- Probability of assignment to treatment does not remain constant throughout trial

- Instead, subject assignment to treatment is determined by current balance across arms of
  - Important prognostic variables (to ensure balance of potential confounders)
  - Outcome (to ensure patients have a greater probability of being randomized to more efficacious treatment)

before random assignment and were repeated every 6 weeks for the first year after random assignment and every 12 weeks thereafter until disease progression. The same radiologic method used to document disease at baseline was used at subsequent assessments. RECIST guidelines were used to define all responses.[23] No independent radiologic review was performed. Survival status was assessed every 3 months after completion of study treatment.

Safety assessments were performed until 28 days after the last exposure to study treatment, followed by an additional 6-month safety follow-up period. Intensity of adverse events was graded according to National Cancer Institute Common Terminology Criteria for Adverse Events, version 3.0. An independent data safety monitoring board regularly reviewed study safety and efficacy data.

Statistical Analysis

The intention-to-treat patient population, the primary population for efficacy analysis, included all randomly assigned patients. The safety population included all randomly assigned patients who received at least one dose of study medication. The measurable disease population was used to evaluate response rate only. In the safety analysis, patients were analyzed as treated. The primary study end point was OS, defined as time between random assignment and death irrespective of cause. Secondary end points were progression-free survival (PFS; defined as time between random assignment and first documented disease progression or death), overall response rate, and safety.

AVAGAST was designed as a group sequential study with up to two data looks, with the final analysis planned after approximately 517 deaths had occurred. Per protocol, the preplanned interim analysis (after two thirds of the expected events [ie, 345]) was dropped because, at the time the analysis was due, it was estimated that the final analysis would follow in ≤ 6 months. On the basis of a systematic literature review, it was assumed that median OS in the placebo group would be 10 months. The study was powered to test the hypothesis that the addition of bevacizumab would improve median OS to 12.8 months, equivalent to a hazard ratio (HR) of 0.78 between study groups, assuming an exponential distribution for the time-to-death variable. Because no interim analysis was performed, the study became a fixed-sample study. To detect an HR of 0.78, 509 deaths were necessary to

Table 1. Baseline Characteristics (intention-to-treat population)

| Variable | Fluoropyrimidine-Cisplatin + | | | |
| | Bevacizumab (n = 387) | | Placebo (n = 387) | |
	No.	%	No.	%
Sex				
Male	257	66	258	67
Female	130	34	129	33
Age, years				
Median	58		59	
Range	22-81		22-82	
ECOG performance status				
0-1	365	94	367	95
≥ 2	22	6	20	5
Geographic region				
Asia-Pacific	188	49	188	49
Europe	125	32	124	32
Pan-America	74	19	75	19
Fluoropyrimidine treatment				
Capecitabine	364	94	365	94
Fluorouracil	23	6	22	6
Primary tumor site				
Stomach	333	86	338	87
Gastroesophageal junction	54	14	49	13
Measurable disease	311	80	297	77
Extent of disease				
Metastatic	367	95	378	98
Locally advanced	20	5	9	2
Liver metastases	130	34	126	33
Previous treatment				
Neoadjuvant therapy	30	8	30	8
Gastrectomy	110	28	107	28

Abbreviation: ECOG, Eastern Cooperative Oncology Group.

Blinding

- *Blinding* = treatment masking
- Treatment masked from the patient - single blind
- Treatment masked from both the patient and the study personnel - double blind
- Blinding controls for
  - Treatment/procedure bias
  - Assessment bias
- No blinding for members of DSMB
- Blinding isn't always possible
  - Drugs being compared have different modes of delivery - infusion versus tablet
  - Blinding in trials of devices or surgical procedures is difficult or impossible

Study populations

Intention-to-treat (ITT) is the idea that patients in a randomized clinical trial should be analyzed as part of the treatment group to which they were assigned, even if they did not actually receive the intended treatment. For assessing efficacy, analysis of the ITT population is preferred.

Treatment received (TR) is the idea that patients should be analyzed according to the treatment actually given, even if the randomization calls for something else. For assessing safety, analysis of the TR population is preferred. This is sensible since we want to attribute any severe adverse events to the treatment actually received.

Efficacy based on the ITT population

- Many factors contribute to a patient's inability to complete the intended therapy or patient adherence
  - Side effects
  - Disease progression
  - Patient/physician preference for a different treatment
- Failure to complete therapy as randomized is almost always an outcome of the study itself
- If ITT population is not analyzed, can result in post-entry selection bias
- Results from efficacy analysis using ITT population is a test of treatment policy/program effectiveness

Planned early data looks

- Early stopping for efficacy
  - Is there sufficient evidence to conclude treatment is significantly better?
  - Ethical imperative not to continue to randomize patients to a therapy known to be inferior
  - Inflates type I error rate
- Early stopping for futility
  - No hope of being able to demonstrate treatment efficacy
  - Sufficient data to answer scientific question
  - Unethical to continue to randomize patients in a trial with no additional benefit
  - Inflates type II error rate (decreases power)
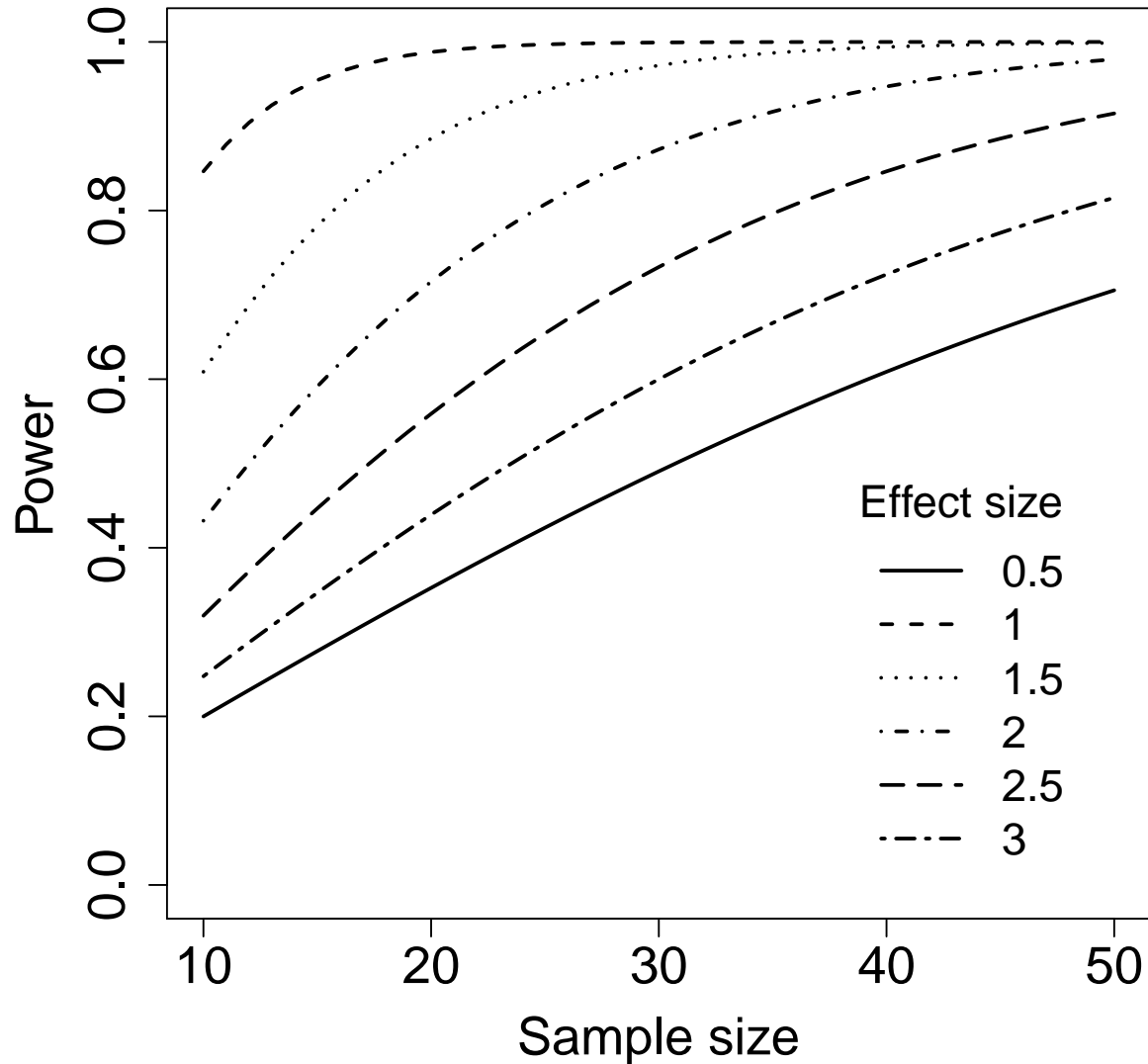- Early stopping for safety

Sample size and power

Power as a function of sample size for increasing SD
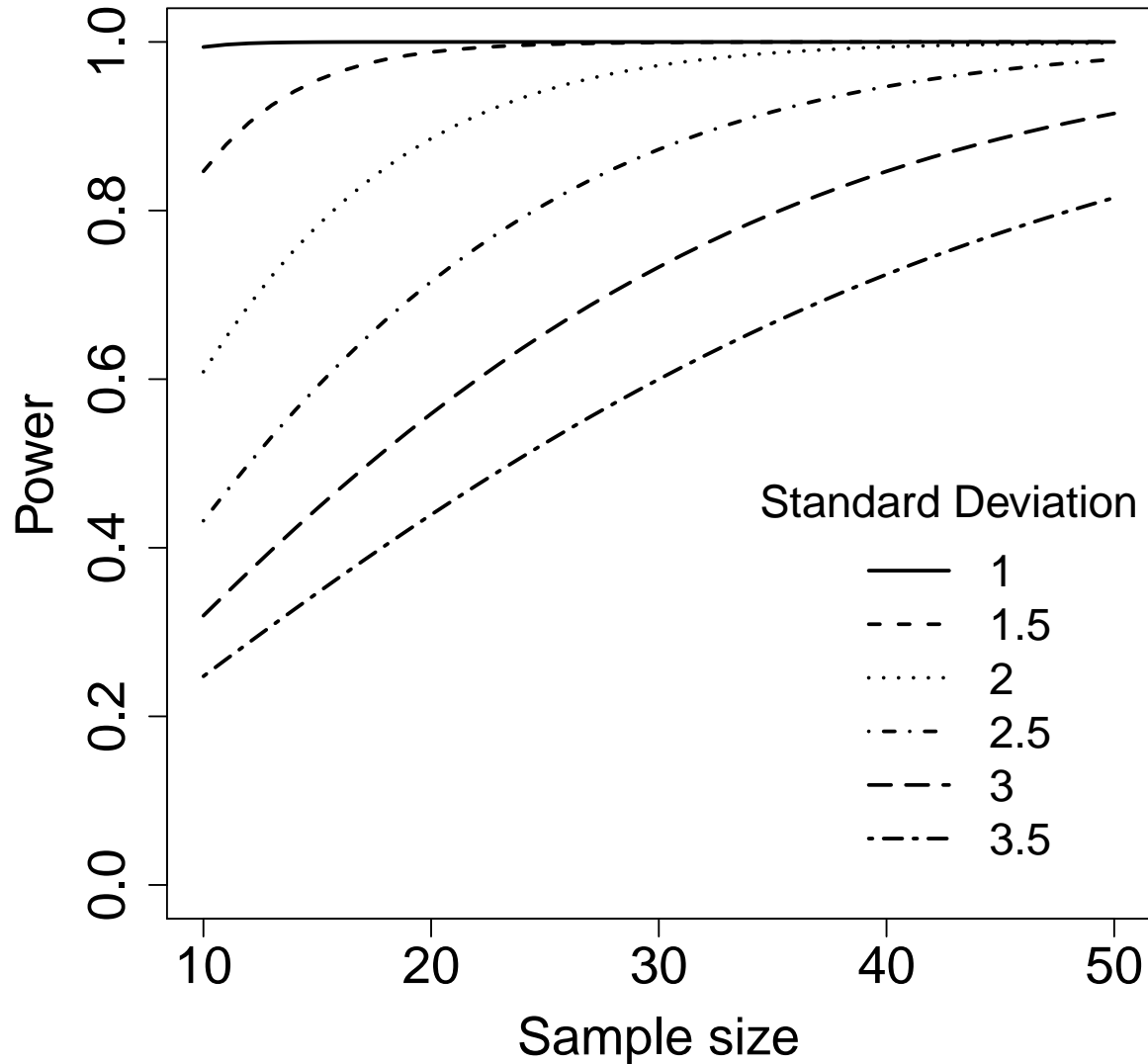
Sample size and power (cont.)

Power as a function of sample size for increasing effect size

Sample size and power (cont.)

Power as a function of sample size for increasing type I error rate