MEDICAL COLLEGE OF WISCONSIN

Designing Clinical Trials

Brent R. Logan, PhD

Sponsored by the Clinical and Translational Science Institute (CTSI)
and the Department of Population Health / Division of Biostatistics

Speaker Disclosure

In accordance with the ACCME policy on speaker disclosure, the speaker and planners who are in a position to control the educational activity of this program were asked to disclose all relevant financial relationships with any commercial interest to the audience. The speaker and program planners have no relationships to disclose.

Outline

- *Overview of Clinical Trial phases*
- *Phase I Trials*
- *Phase II Trials*
- *Phase III Trials*
  - *Endpoints*
  - *Eligibility*
  - *Randomization*
  - *Blinding*
  - *Power and Sample Size*
  - *Data Monitoring*
  - *Alternative Study Designs*

Phases of Human Research

- *Ideal clinical trial: randomized, double-blinded*
- *Traditionally four phases of research*
  - *I=Establish safety, dose-finding, PK studies*
  - *II=Establish biological activity or potential efficacy*
  - *III=Randomized comparison of treatment*
  - *IV=Long-term surveillance in broader population*
  - *Some hybrid phases (I/II, II/III)*

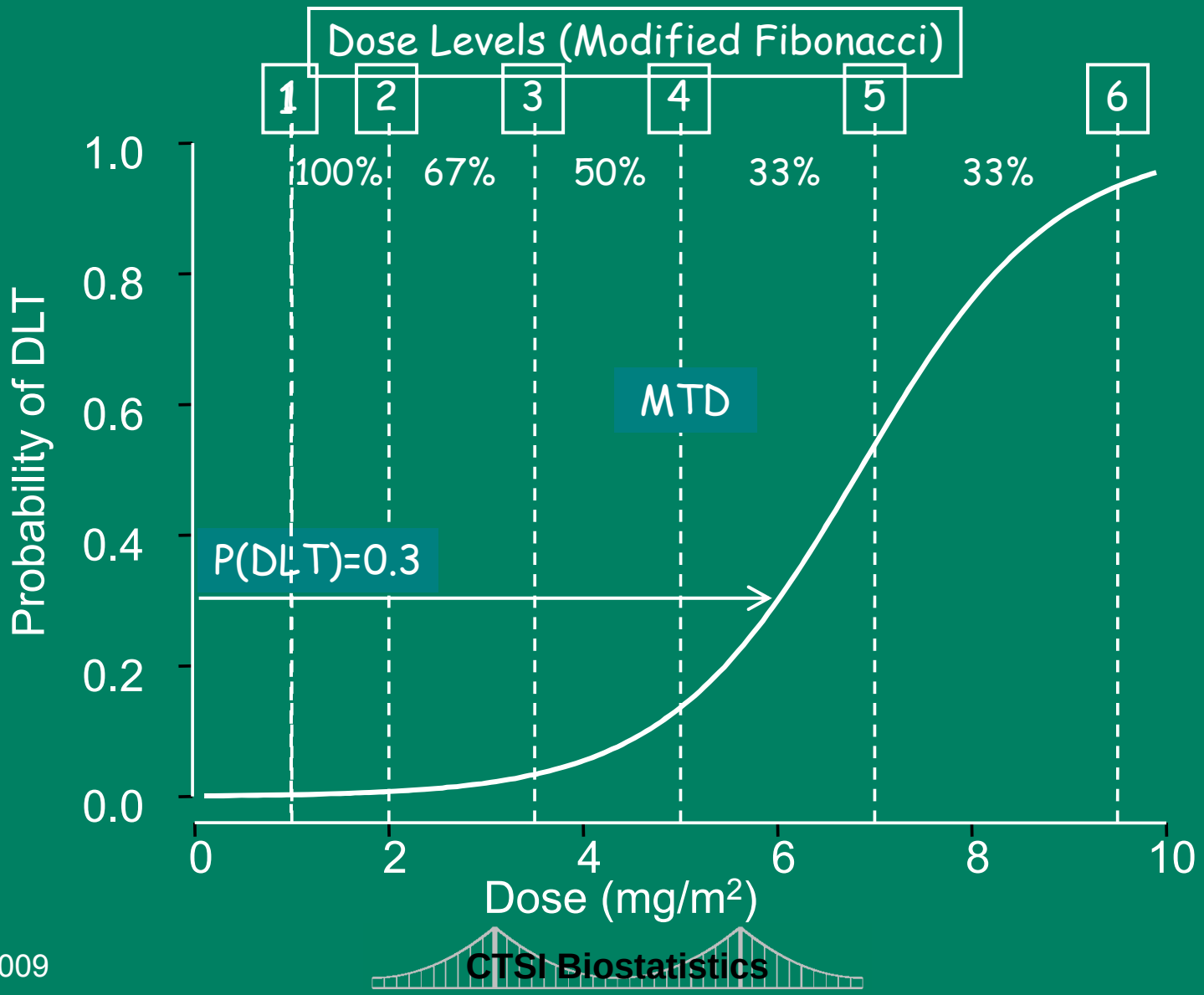
Phase I Trials

- Goal:
  - How well is the drug tolerated in small numbers of individuals?
  - Toxicity screening and determination of maximum tolerated dose (MTD)
- Phase IA – Healthy subjects
- Phase IB – Diseased subjects, failures on conventional therapy
- Small studies - < 10 per dose
- Dose escalation or de-escalation plan
- Efficacy measured but not the main objective

Phase I Trials

- ***Maximum Tolerated Dose:***
  - ***Highest possible dose with an "acceptable" rate of dose-limiting toxicities***
  - ***"acceptable" toxicity rate set by investigators because of potential for benefit to patients***
  - ***Example:***
    - ***Dose which produces grade III or worse toxicities in no more than 1 in 3 patients (cancer studies)***

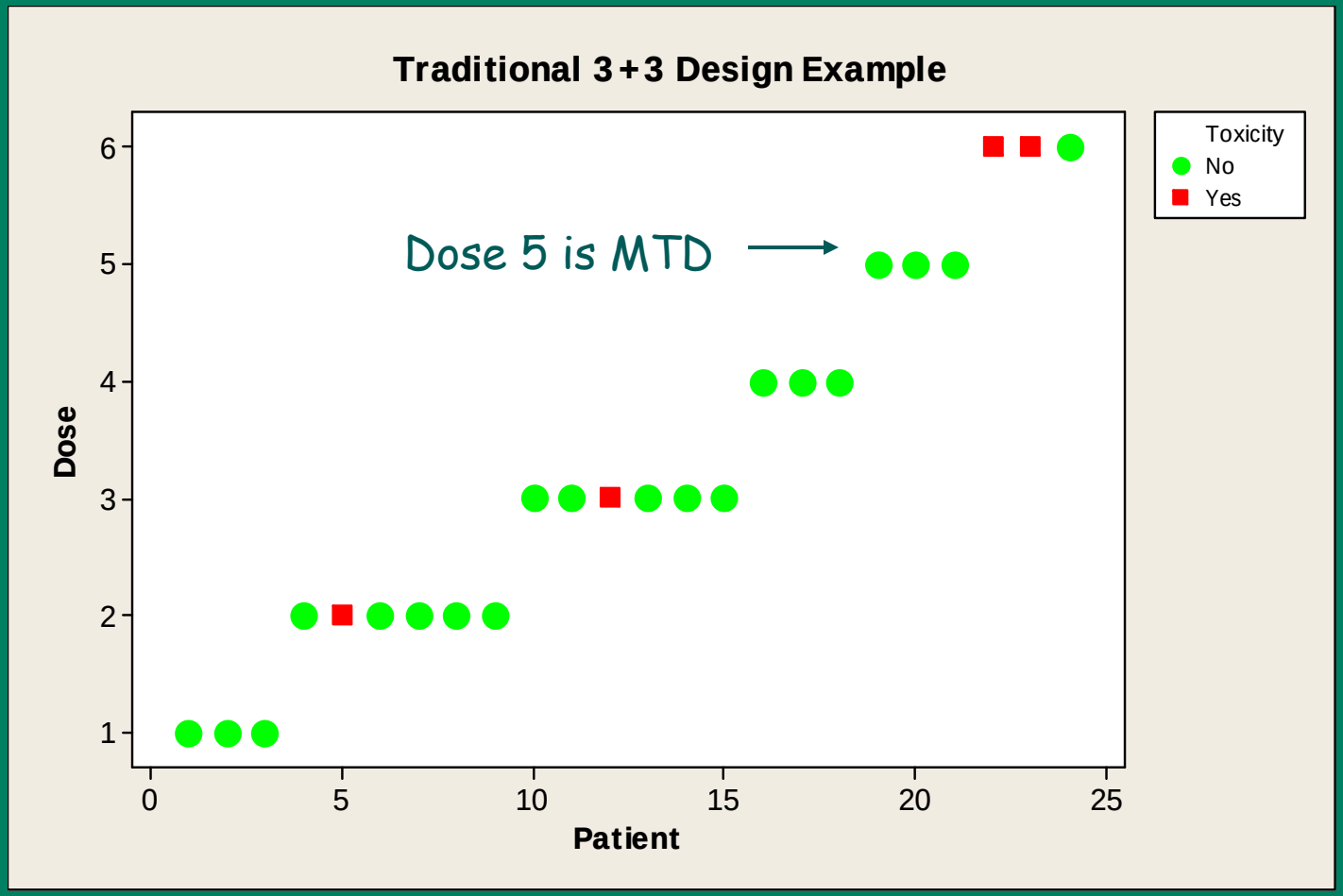
Dose Response



Traditional 3+3 Design

- ***Patients enrolled in groups of 3 for each dose level.***
  - ***None of 3 experience DLT– increase dose level***
  - ***1 of 3 experience DLT - enroll additional 3 patients***
  - ***>1 of 3 experience DLT – stop; MTD was reached at previous dose***
  - ***If <2 out of 6 experience DLT then increase dose level***
  - ***If 2 or more out of 6 experience DLT, then MTD was reached at previous dose.***

Traditional 3+3 Design



Traditional 3+3 Design

- *Drawbacks of 3+3 design*
  - *Biased estimate of MTD (too low)*
  - *MTD defined based on 33% DLT rate*
  - *Moves slowly through the doses*
- *Advantages:*
  - *Simple: No statistical modeling needed*

Other Phase I Designs

- *Continual Reassessment Method (O'Quigley, Biometrics 1990):*
  - *After each patient's outcome is known, re-fit the dose-toxicity curve*
  - *Estimate MTD based on the estimated curve*
  - *Enroll next patient at estimated MTD.*
  - *Many modifications to this principle*
  - *More flexible, better statistical properties than 3+3 design*
  - *Requires collaboration with statistician*

Phase II Trials

- *Goal: Determination if treatment has any biological activity, estimation of biological activity, estimation of rate of adverse events*
- *Single well described treatment regimen*
- *Small sample sizes - <40-50*
- *Used to select treatment for Phase III trials*
- *Tighter inclusion/exclusion criteria than Phase III trial – who will benefit most?*
- *"Quicker" outcome measures if possible*
- *Often two-stage design*

Phase II Trials: Goals

- *Estimation of outcome*
  - *Construct confidence intervals for outcome*
  - *Outcome often binary (yes/no) with success probability p*
- *Establishing efficacy*
  - *Often uses historical control success rate (p_0)*
  - *Declare effective if Hypothesis test rejects H_0:$p=p_0$*
  - *Equivalently, if confidence interval for p does not contain p_0.*
  - *Significance level alpha = probability of declaring treatment promising when $p=p_0$*

Phase II Trial Example

- *Augmenting high-dose chemotherapy with Interleukin-2 activated autotransplant, for metastatic breast carcinoma (Toh BMT 2000)*
- *33 patients with de novo or relapsed stage I to III breast cancer between 1996-1998.*
  - *2 yr. PFS=35% (95% CI of 15-54%)*
  - *2 yr. OS=78% (95% CI of 62-94%).*
  - *Collected toxicity data – 2 grade III toxicities*
  - *Collected engraftment data – all engrafted*
- *Historical control – 29 patients between 1993 and 1997*
  - *2 yr. PFS=17%, 2 yr. OS=61%. Not sig. Diff.*

Phase II Trials: Sample Size

- *Precision of estimation:*
  - *larger n yields narrower confidence interval*
  - *Specify targeted width of CI or margin of error (half-width)*
- *Power: Chance of declaring treatment effective when $p=p_1$*
  - *Larger n yields greater power*
  - *Specify p_0, p_1, significance level, power*
- *Example: Steroids for grade II-IV acute GVHD: historically CR rate=35%*
  - *New agent: Want 80% power to detect improvement in CR rate to 55%. With one-sided test, alpha=0.05, n=35*

Two-Stage Designs

- *Want to minimize the number of patients treated with an ineffective treatment*
- *Choose n_1 patients in first stage*
  - *If r_1 or fewer respond then declare the treatment a failure and stop*
  - *If $>r_1$ respond then add $n-n_1$ patients (total=n)*
- *If more than r respond out of n then declare the treatment a success; o/w declare it a failure*
- *Requires short-term primary endpoint*

Two-Stage Design

- ***Example: Suppose we want to compare a new treatment against a historical control with response rate 0.1. Want 80% power when response rate is 0.3, with α=0.05***
- ***EN=Expected sample size***
- ***PET=Probability of Early Termination***

		Reject drug if			
p_0	p_1	$\leq r_1/n_1$	$\leq r/n$	$EN(p_0)$	$PET(p_0)$
0.1	0.3	1/10	5/29	15.0	0.74

Two-Stage Designs

- *Simon (Cancer Treatment Reports, 1985)*
- *Calculator for these designs can be obtained online at http://linus.nci.nih.gov/brb/samplesize/otsd.html*

Phase III Trials: Design Features

- *Testable hypothesis*
  - *Outcomes*
  - *Patient population*
- *Design*
  - *Control group*
  - *Randomization*
  - *Blinding*
  - *Data Monitoring*

What is the Question?

- *Includes*
  - *Treatment Comparison*
  - *Primary Outcome*
  - *Patient Population*
- *Usually one primary question; clearly defined and stated in protocol*
- *Secondary questions*

Primary Endpoint

- *Main clinical variable of interest which is most likely to be impacted by treatment*
- *Could be composite endpoint, e.g. for studying heart disease, either*
  - *Death from coronary heart disease*
  - *Nonfatal myocardial infarction*
- *Response variables should be capable of unbiased assessment*
  - *Blinding*
- *Important to have response variables which can be ascertained as completely as possible*
  - *Loss to follow-up, missed appointments*

Primary Endpoint Example

- *HIV studies: HIV virus destroys immune system, individuals subject to infections which lead to death; now treatments target the virus or specific infections*
- *Possible endpoints:*
  1. *Increase in CD4 counts (direct measure of immune function)*
  2. *Viral RNA reduction (measures amount of virus in the body)*
  3. *Time to one or more opportunistic infections*
  4. *Time to death from any cause*
  5. *Time to first infection or death*

Primary Endpoint Example

- *1,2 may be appropriate for phase II trial*
- *3 ignores competing risk of death without infection*
- *4 is of ultimate interest, but may not be practical as mortality rate is slowed down and patients have a long expected survival*
- *5 is often used, but the definition of opportunistic infection must be clear*
- *Many of these are important for secondary analysis*

Sometimes secondary endpoints can be important

Secondary Questions

- *Components of composite endpoint*
- *Other endpoints*
- *Subgroup analyses*
- *Toxicity*
- *Ancillary study questions*
- *Surrogate outcomes*
- *Multiple Testing:*
  - *If enough tests are done, some will be significant by chance alone when there is no intervention effect*
- *Secondary hypotheses are better at shedding light on results, or on generating new hypotheses; should not be considered definitive results*

Subgroup Analysis

- *Trial should have reasonable expectations that intervention will be consistent across subgroups; otherwise, do separate studies*
- *Is effect of treatment in a clinical trial homogeneous across all patients in that trial?*
- *Example 1: ISIS-2 Trial: 17000 patients with AMI randomized to placebo vs. aspirin (also streptokinase) (Lancet, 1988)*
  - *Mortality within 1 month: 9% (aspirin) vs. 12% (placebo), p<0.001*
  - *Investigators were urged (by editors) to conduct nearly 40 subgroup analyses*
  - *Investigators agreed on condition that they could conduct their own subgroup analysis to illustrate unreliability of subgroup findings*

Subgroup Analyses

- Subgroup defined by astrological sign

	# of deaths in 1 month		
Astrological sign	Aspirin	Placebo	p-value
Libra or Gemini	150	147	NS
Others	654	869	<0.001

- Increased variability of results just due to chance when you look at a lot of subgroups.
- Excess of type II errors due to multiple comparisons

Primary Question: Example

- *Thompson (Blood, 2008) SWOG 9438*
- *Treatment:*
  - *Post-transplant therapy with IL-2 versus observation after TBI, etoposide, cyclophosphamide and PBSC autotransplant*
- *Patient Population:*
  - *Relapsed NHL*
- *Outcome: Progression-free survival*

Secondary Questions: SWOG NHL Example

- ***Secondary Endpoints***
  - ***OS, Toxicity***
- ***Subgroup analyses***
  - ***Test for interaction between chemosensitivity and outcome, or between histology and outcome***
  - ***No significant interactions***

Patient Population

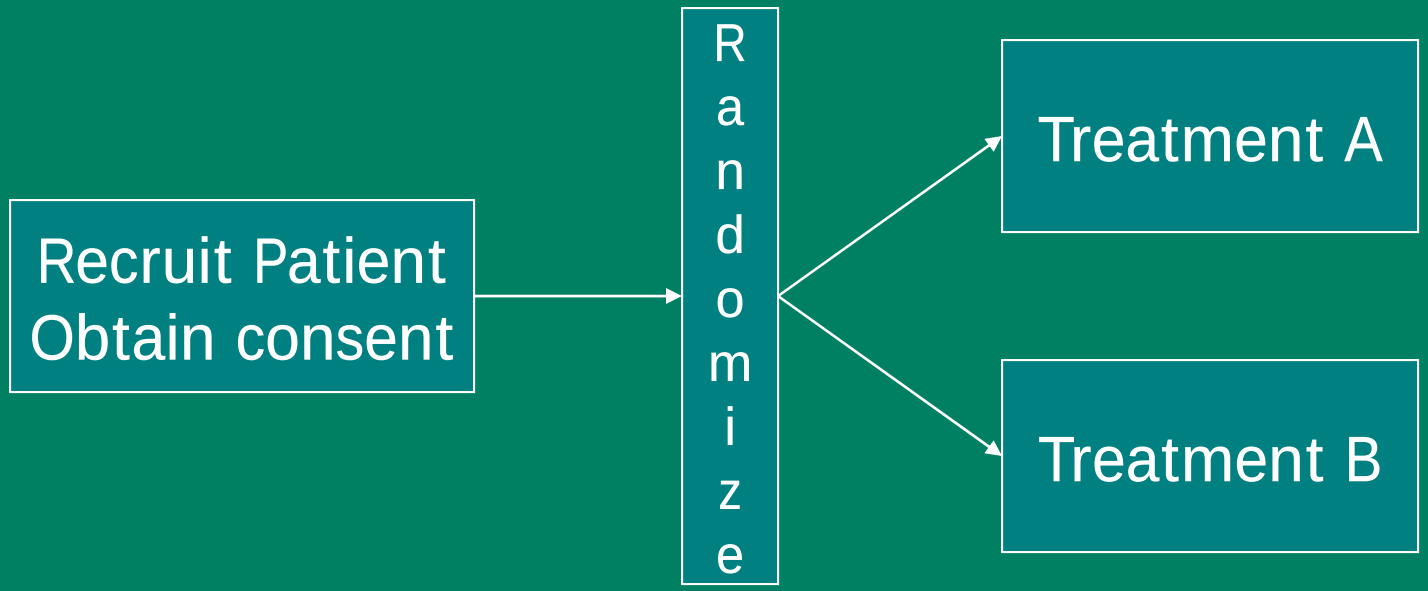
- *The study population should be defined in advance using clearly defined inclusion/exclusion criteria*
- *Phase II vs. phase III*
- *Reproducibility of results*
  - *To whom do the results apply?*
  - *Allow other investigators to assess merits*
- *Applicability of inference to future patients*
  - *Patients volunteering for trials are often different than general patient population*
- *Recruitment: Homogeneity vs. Feasibility*

Patient Population

- *Selecting the patient population*
  - *Potential or high likelihood to benefit*
    - *High event rate*
    - *Ex: Anti-anginal drug, don't enroll patient with only one angina episode in last 2 years*
  - *Weigh benefits against adverse effects*
    - *Pregnant women*
    - *Gastric bleeding for anti-inflammatory drugs*
  - *Low chance of competing risks*
    - *For studies of heart disease, exclude patients with cancer*
  - *Likely to adhere to study protocol*

Randomized Phase III Design Schema

- ***Each patient randomized to one of two treatments:***



Randomization

- *Process by which all participants are equally likely to be assigned to either the intervention or control groups*
- *Advantages:*
  - *Protects against bias in treatment allocation and selection of patients*
  - *Makes groups comparable or "similar on the average", balancing out known and unknown prognostic factors*
    - *An individual study may still be imbalanced*

Selection Bias with Nonrandomized Controls

- ***Pilot trial of high-dose chemotherapy with autotransplant for multiple myeloma (Barlogie 1995) – 3 yr. Survival = 60%.***
  - ***Historical control: SWOG series of four Phase III trials of standard chemotherapy between 1977 and 1989 – 3 yr. Survival = 40%.***
  - ***Bias in historical comparison due to different eligibility criteria – transplant patients must be young and in good condition.***
  - ***When SWOG patients restricted to <70 yrs old with good renal function, 3 yr. Survival=60%.***

Methods of Randomization

- *Simple randomization (Coin flip)*
  - *Allows for unusual results in treatment assignment*
- *Random permuted blocks ensure balance over accrual time*
  - *ABBA | BABA | BBAA | ABAB …*
- *Stratification ensures balance across important covariates known to be associated with outcome*
  - *Often stratification by center*

Randomization Issues

- ***When to randomize: As close to treatment as possible to minimize dropout.***
- ***Intention to Treat Principle: All eligible patients are analyzed according to the arm to which they are randomized, even if they deviate from the therapy specified in the protocol***
- ***Avoids bias – e.g. if good-risk patients decide toxicity is not worth small expected benefit from treatment***

Randomization and ITT: Example

- ***SWOG NHL trial of IL-2 post tx***
  - ***When should randomization be done?***

Example (cont.)

- *Registration prior to tx*
- *Randomization took place between day 28 and day 80 post tx*
- *Additional eligibility criteria regarding patient's status at time of randomization (no recurrence/ progression of disease, no serious toxicity)*
- *394 patients initially registered, 376 eligible*
- *194 were randomized*
- *182 eligible but no randomization*
  - *46 Refusal*
  - *62 grade 4 or 5 toxicity*
  - *28 progression*

Design Issues-Blinding

Design Issues-Blinding

- *Why patient?*
  - *May experience a psychological benefit of being on new treatment*
  - *Attitude toward treatment may affect cooperation in study (compliance with therapy, attendance for evaluation, drop out rates)*
  - *Attitude toward treatment may affect response*

Design Issues - Blinding

- *Why Treatment team?*
  - *Knowledge of treatment may influence:*
    - *Decisions on dose modification*
    - *Decisions on intensity of patient examination*
    - *Decisions on when to discontinue treatment*
    - *Decisions on ancillary care*
  - *Enthusiasm for treatment may be expressed to patient*

Design Issues-Blinding

- *Why evaluator?*
  - *Objectivity of evaluator may be lost*
  - *Leads to assessment bias*
- *Is blinding appropriate?*
  - *Ethical issues: No undue harm to patient*
  - *Practical issues:*
    - *Need treatments similar (Dose schedules, similar pills)*
    - *Informed consent is harder*
    - *Administration is harder*

Sample Size

- *Patients respond differently to treatment; we are interested in collective effect of treatment*
- *Larger sample sizes increase precision of estimated treatment effect.*
- *Calculations based on primary endpoint: Specify likelihood of detecting a significant treatment effect of a particular magnitude (Power)*
- *Example: Want 80% power to detect if new treatment increases 2 year progression free survival from 40% to 55%*

Sample Size

- *Need to specify:*
  - *Type I error*
  - *Power*
  - *Success proportion for control (p_1) and intervention (p_2)*
- *Formula: N per group is*

$$N = 2(z_\alpha + z_\beta)^2 \, p_c(1-p_c)/(p_2-p_1)^2$$

- *Here*
  - *z_α=1.96 for a two-sided 5% level test*
  - *z_β=1.28 and 0.84 for 90% and 80% power*
  - *p_c is the average success proportion*

Sample Size

- *Tables*
  - *Hulley et al. (1988)*
  - *Example: For P_1=0.35,P_2=0.55,Power=80%, type I error=5%,*
    - *n=96 per arm*
- *Free software/calculators on Internet*

  http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize
  http://www.math.uiowa.edu/~rlenth/Power/

- *Other methods for different endpoints*
  - *Continuous Normal outcomes*
  - *Survival outcomes*
- *Consult your statistician*

Data Monitoring

- *Data Safety Monitoring Committee: Responsible for ensuring patient safety and study integrity*
- *Meets every 6-12 months to review data*
- *Justifications for early termination of study*
  - *Unacceptable toxicity*
  - *Accrual problems*
  - *Efficacy or Futility if included in design*

Data Monitoring

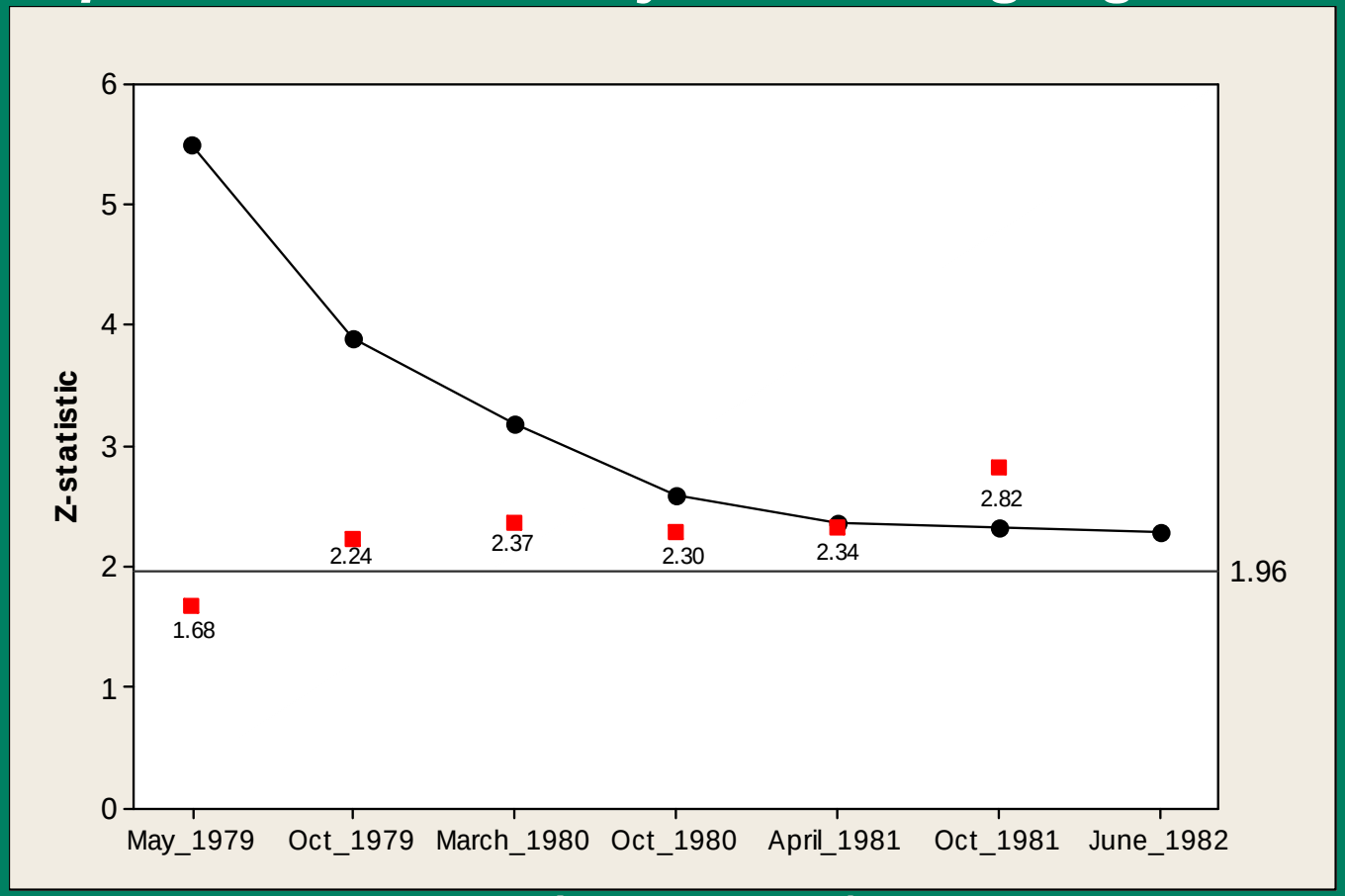
- *Often specify toxicity stopping rules in protocol*
- *Trigger DSMB review*
- *Short-term outcome*
  - *Grade III-V toxicities within 1 month*
- *Important to understand expected toxicity rates*
  - *Compare observed rates to expected rates*

Group Sequential Designs

- *Decision on trial is made sequentially based on groups of patients entered in the study.*
- *Reasons for termination of trial at interim evaluation:*
  - *Greater than expected benefit of treatment*
  - *Worse than expected effect of treatment-toxicity*
  - *Futility – little or no chance of obtaining a statistically significant result by end of study*
- *Stopping early generally requires short-term endpoint*
- *Adjustment made to significance level at each interim analysis*

Beta Blocker Heart attack trial (DeMets CCT 1984)

- *Comparison of mortality rates using log-rank test*



Basic Data Items and Forms

- *Detailed in protocol*
- *Basic variables:*
  - *Eligibility*
  - *Evaluability*
  - *Treatment Summary*
  - *Outcome Summary*
- *Forms*
  - *Study-specific flow sheet*
  - *Standard off-treatment notice/notice of death*
  - *Prestudy form (baseline characteristics)*

Sample Protocol (Friedman et al. 1998)

1. Background of the Study
2. Objectives
    1. Primary question and response variable
    2. Secondary questions and response variables
    3. Subgroup hypotheses
    4. Adverse effects
3. Design of the Study
    1. Study Population-Inclusion/Exclusion criteria
    2. Sample size assumptions and estimates
    3. Enrollment of participants
        1. Informed consent
        2. Assessment of eligibility
        3. Baseline examination
        4. Intervention allocation (randomization)

Sample Protocol

4. Intervention
    1. Description and Schedule
    2. Measures of compliance
    3. Dose modification and withdrawals
5. Follow-up visit description and schedule
6. Ascertainment of response variables
    1. Data collection and forms
    2. Quality control
7. Data Analysis
    1. Interim Monitoring
    2. Final analysis
8. Termination Policy

4. Administrative organization

Resources

- The Clinical and Translation Science Institute (CTSI) supports education, collaboration, and research in clinical and translational science: www.ctsi.mcw.edu
- The Biostatistics Consulting Service provides comprehensive statistical support

  http://www.mcw.edu/biostatsconsult.htm

Free drop-in consulting

- MCW/Froedtert/CHW:
  - Monday, Wednesday, Friday 1 – 3 PM @ Froedtert Pavilion, Room #L777A (TRU Offices)
  - Tuesday, Thursday 1 – 3 PM @ Health Research Center, H2400
- VA: 1st and 3rd Monday, 8:30-11:30 am
  - VA Medical Center, Building 111-B-5423
- Marquette: 2nd and 4th Monday, 8:30-11:30 am
  - Olin Engineering Building, Room 338D

Upcoming Lectures

Wednesday, Sept 30, 2009 at 8:50 AM
(Jennifer Le-Rademacher, PhD)
Statistics, Probability and Diagnostic Medicine
Conference Room M-3rd Floor of the Clinical Cancer Center

Thursday, October 22, 2009 at 1:30 PM
(Jennifer Le-Rademacher, PhD)
Statistics, Probability and Diagnostic Medicine
Children's Hospital Auditorium

Friday, October 9, 2009 at 7 AM
(Dan Eastwood, MS and Emily McGinley, MS)
Getting Help for your Bioethics, Biostatistics, and Epidemiology Questions: Population Health Consulting Services
Froedtert NT2209

Wednesday, November 18, 2009 at 8:50 AM
(Brent Logan, PhD)
Designing Clinical Trials
Location: TBA

For locations that are TBA please check the website below two weeks prior to the lecture date:
http://www.mcw.edu/biostatistics/CalendarCurrentEvents/SeminarSeries.htm