Biost 578A: Special Topics
Statistical Design of Clinical Trials

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

Lecture 1:

Course Structure; Overview

April 1, 2013

1

Lecture Outline

- Course objectives
- Course structure
- Overview of setting
  - Elements of RCT design
  - Constraints and optimality criteria

2

Course Objectives

3

Learning Objectives

- Review of scientific, ethical, logistical constraints in RCT setting
- Application of standard statistical theory to RCT setting
- Exposure to statistical methods more specific to RCT
- Elements of complete statistical design of an RCT

4

Constraints in RCT Setting

- Human experimentation to evaluate treatments
  - Scientific rigor
  - Individual and group ethics
    - Clinical science vs basic science
  - Regulatory environment
    - Game theory
  - Logistical issues
    - Feasibility
    - Cost of RCT
    - Return on investment

5

Standard Theory in RCT Applications

- Parametric vs semiparametric vs distribution-free models
- Frequentist and Bayesian paradigms for inference
- Asymptotic methods for inference
- Censored data methods
- Clustered data methods
- Regression methods
- Relative efficiency and asymptotic relative efficiency
- Multiple comparison issues

6

RCT Methodology

- Randomization strategies
  - CRD; blocking; stratification; dynamic balancing; response adaptive randomization
- Sequential methods
  - Group sequential; adaptive designs; Bayesian methods
- Noninferiority trials
- Missing data methods in RCT setting

7

Complete Statistical Design of RCT

- Grant application
- Protocol
- (ClinicalTrials.gov)
- (Case Report Forms (CRFs))
- (Manual of Operations)
- Statistical analysis plan
- Interim monitoring plan
- Clinical trial report
- Publications / presentations in scientific literature

8

Warning

- From the McCarthy Era:

Politician: Are you still teaching communism at the University of Chicago?

President: Yes, and in the medical school, we are still teaching cancer.

9

Disclaimer

- From the "Prologue – Concerning Boggies":

"This book is predominantly concerned with making money, and from its pages a reader may learn much about the character and the literary integrity of the authors. Of boggies, however, he will discover next to nothing, since anyone in the possession of a mere moiety of his marbles will readily concede that such creatures could exist only in the minds of children of the sort whose childhoods are spend in wicker baskets, and who grow up to be muggers, dog thieves, and insurance salesmen."

Harvard Lampoon, *Bored of the Rings*

10

Course Structure

11

Lectures

- Instructor: Scott S. Emerson, M.D., Ph.D.
  - Office hours by appointment
- Time and Place:
  - Lectures: 8:00 - 9:20 am MW HST 663
  - (Lectures on May 20 and May 22 will be rescheduled)

12

Course Web Pages

- Address: www.emersonstatistics.com/b578/
- Content
  - Syllabus
  - Lecture handouts (eventually)
  - Links to www.RCTdesign.org for
    - Recordings of lectures (eventually)
    - Occasional white papers
    - Examples of code

13

Assumed Prior Knowledge

- Statistical coursework
  - Introductory Applied Biostatistics (Biost 514 – 515)
  - Mathematical Statistics (Stat 512 – 513)
  - Scientific Design of Medical Studies (Biost 524)

14

References

- Jennison & Turnbull (2000)
  - *Group Sequential Methods with Applications to Clinical Trials* (Chapman and Hall / CRC)
- Whitehead (1997: 2nd Rev Ed.)
  - The Design and Analysis of Sequential Clinical Trials (Ellis Horwood)
- Proschan, Lan & Wittes (2006)
  - Statistical Monitoring of Clinical Trials: A Unified Approach (Springer)
- www.RCTdesign.org
  - Technical reports, tutorials, RCTdesign documentation

15

Computer Software

- Students may use any program that will do what is required,
- RCTdesign is a free R package (Emerson)
  - License from www.RCTdesign.org
- East (Mehta, Pampallona)
  - www.cytel.com
- PESTv4 (Whitehead)
  - www.mps-research.com/PEST
- SAS SEQDESIGN and SEQTEST

16

Weekly Homeworks

- Typically assigned Wed due at midnight on Sunday
- Usually involve derivations or simulations to:
  - Illustrate justification for standard RCT practices
  - Quantify advantages / disadvantages of particular approaches
  - Extend methods covered in class

17

Presentation

- Each student will make a 20 minute presentation to the class on a selected topic
- I will provide a list of possible topics, but other topics are possible with my approval

18

Grading

- Based on performance on homeworks and presentation
- CR / NC

19

Overview of Setting

Elements of RCT Design: Overall Goal

20

Clinical Trials

- Experimentation in human volunteers
- Investigates a new treatment/preventive agent
  - Safety:
    - Are there adverse effects that clearly outweigh any potential benefit?
  - Efficacy:
    - Can the treatment alter the disease process in a beneficial way?
  - Effectiveness:
    - Would adoption of the treatment as a standard affect morbidity / mortality in the population?

21

Overall Goal

- "Drug discovery"
  - More generally
    - a therapy / preventive strategy or diagnostic / prognostic procedure
    - for some disease
    - in some population of patients
- A series of experiments to establish
  - Safety of investigations / dose
  - Safety of therapy
  - Measures of efficacy
    - Treatment, population, and outcomes
  - Confirmation of efficacy
  - Confirmation of effectiveness

22

The Enemy

"Let's start at the very beginning, a very good place to start…"

- Maria von Trapp

(as quoted by Rodgers and Hammerstein)

23

First: Where do we want to be?

- Describe some innovative experiment?
- Find a use for some proprietary drug / biologic / device?
  - "Obtain a significant p value"
- Find a new strategy that improves health of some individuals
  - "Efficacy" seems to benefit some people somehow
- Find a new strategy that improves health of the population
  - "Effectiveness" requires proper use of a strategy that modifies an important clinical endpoint

24

6

Efficacy vs Effectiveness

- An efficacious treatment has demonstrated an ability to beneficially modify
  - An endpoint thought to be an indicator of good clinical outcome
  - In some subset of patients
  - Under some conditions that are at least marginally relevant
- An effective treatment is one whose introduction improves health in the population

25

Effectiveness: A Moving Target

- A treatment can be both efficacious and ineffective depending on factors of clinical trials and changes in clinical practice
- Target population
  - Rigor of diagnosis, severity, eligibility criteria, run-ins, off-label use, change in behaviors
- Control treatment
  - Proscribed and prohibited ancillary treatments
- Intervention
  - Adherence
- Measurement of outcome(s)
  - Clinical vs subclinical, timeframe
- Summary measure of outcome distribution

26

Ex: Screening for Lung Cancer

- RCT of spiral CT versus chest X-ray (cf: NLST)
- Efficacy trial might consider
  - High risk, asymptomatic smokers
  - Three annual screens by highly trained radiologists
  - Protocol defined follow-up screening / treatment
  - Outcome is cancer diagnosis or mortality due to lung cancer
- Effectiveness might depend on
  - Broader criteria for use
  - Annual screening by community radiologists
  - Follow-up per local medical standards
  - Changes in underlying risk factors or concomitant disease
  - Outcome is all cause mortality

27

Ex: Treatment for Lung Cancer

- RCT of experimental agent vs standard of care
- Efficacy trial might consider
  - Patients with measurable disease failing prior chemotherapy
  - Protocol defined follow-up radiology on frequent basis
  - Outcome is objective response rate or (subclinical) cancer progression free survival
- Effectiveness might depend on
  - Broader criteria for use
  - Follow-up per usual standard of care
    - Ancillary treatments
  - Outcome is all cause mortality

28

Which: Efficacy or Effectiveness

- Factors leading to efficacy trials
  - "Knowledge is good"
  - As pilot studies before prevention studies
  - Inability to perform experiment under realistic conditions
- Factors leading to effectiveness trials
  - Serious conditions
    - Patients generally want to get better
  - Short therapeutic window for treatment
  - Waiver of informed consent
    - Do not withhold beneficial treatments in order to establish mechanisms
  - High cost of clinical trials (time, people, $$)

29

My Perspective

- Public Health is primary goal
  - Improve average health of population
  - Patients willing to participate in RCT are a public resource
- Must be aware of economic reality
  - Most "drug discovery" RCT funded by industry
  - Government funded research for orphan or high risk areas
- Must be aware of conflicts of interest
  - Academic researchers' "pet" hypotheses that might bring fame
  - Industry with proprietary drugs that might bring wealth
  - Political / public pressure on Government funders and regulators

6

30

Treatment "Indication"

- Disease
  - Putative cause vs signs / symptoms
    - May involve method of diagnosis, response to therapies
- Population
  - Restrict by risk of AEs or actual prior experience
- Treatment or treatment strategy
  - Formulation, administration, dose, frequency, duration, ancillary therapies
- Outcome
  - Clinical vs surrogate; timeframe; method of measurement

31

Preventive Strategy "Indication"

- Disease
  - Risk classification
- Population
  - Restrict by contraindications
- Treatment or treatment strategy
  - Formulation, administration, dose, frequency, duration, ancillary therapies
- Outcome
  - Clinical vs surrogate; timeframe; method of measurement

32

Diagnostic / Prognostic "Indication"

- Disease
  - Diagnosis: Risk classification
  - Prognosis: Stage of disease
- Population
  - Restrict by contraindications
- Diagnostic / prognostic strategy
  - Formulation, administration, frequency, duration
- Outcome
  - Clinical vs surrogate; timeframe; method of measurement

33

Evidence Based Medicine

- Decisions about treatments, preventive strategies, and diagnostics should consider PICO
  - Patient (population)
  - Intervention
  - Comparators
  - Outcome
- There is a need for estimates of safety, effect

34

Burden of Proof

- Basic science
  - Understanding biologic pathways, mechanisms of action
- Regulatory approval
  - Is drug safe, efficacious, effective?
- Regulatory postmarketing safety
- Comparative effectiveness
  - Comparisons made among approved therapies
  - Relative effectiveness
  - Cost effectiveness

35

U.S. Regulation of Drugs / Biologics

- Wiley Act (1906)
  - Labeling
- Food, Drug, and Cosmetics Act of 1938
  - Safety
- Kefauver – Harris Amendment (1962)
  - Efficacy / effectiveness
    - " [If] there is a lack of substantial evidence that the drug will have the effect ... shall issue an order refusing to approve the application. "
    - "...The term 'substantial evidence' means evidence consisting of **adequate and well-controlled investigations, including clinical investigations**, by experts qualified by scientific training"
- FDA Amendments Act (2007)
  - Registration of RCTs, Pediatrics, Risk Evaluation and Mitigation Strategies (REMS)

36

Adequate & Well-controlled Studies

- Two important aspects
- Randomized clinical trials rather than observational studies
- Confirmatory clinical trials (note the plural)

37

Why RCT?: Real-life Examples

- Effects of arrhythmias post MI on survival
  - Observational studies: high risk for death
  - CAST: Specific anti-arrhythmics have higher mortality
- Effects of beta-carotene on lung CA and survival
  - Observational studies: high dietary beta carotene has lower cancer incidence and longer survival
  - CARET: beta carotene supplementation in smokers leads to higher lung CA incidence and lower survival
- Effects of hormone therapy on cardiac events
  - Observational studies: HT has lower cardiac morbidity and mortality
  - WHI: HT in post menopausal women leads to higher cardiac mortality

38

Key Point

- Treated populations in the observational studies can be substantially different than those in RCT
  - WHI: peri-menopausal vs post-menopausal
- Treatment in observational studies can be substantially different than those in RCT
  - CARET: dietary vs supplemental beta-carotene
- Outcomes in observational studies can be substantially different than those in RCT
  - CAST: arrhythmias in absence of treatment vs arrhythmias in presence of treatment

39

Multiple Comparisons in Biomedicine

- Observational studies
  - Observe many outcomes
  - Observe many exposures
  - Perform many alternative analyses
    - Summary of outcome distribution, adjustment for covariates
  - Consequently: Many apparent associations
    - May be type I errors
    - But even when valid, may be poorly understood due to confounding
- Interventional experiments
  - Exploratory analyses (“Drug discovery”)
    - Modification of analysis methods
    - Multiple endpoints
    - Restriction to subgroups

40

Statistics and Game Theory

- Multiple comparison issues
  - Type I error for each endpoint – subgroup combination
    - In absence of treatment effect, will still decide a benefit exists with probability, say, .025 in each such combination
- Multiple endpoints and subgroups increase the chance of deciding an ineffective treatment should be adopted
  - This problem exists with either frequentist or Bayesian criteria for evidence
  - The actual inflation of the type I error depends
    - the number of multiple comparisons, and
    - the correlation between the endpoints
- Later: Impact of increased type I error on Bayes factor is huge
  - Ratio of power to type I error means multiplicative effects

41

Confirmatory Trials

- The major goal of a "registrational trial" is to confirm a result observed in some early phase study
- Rigorous science: Well defined confirmatory studies
  - Eligibility criteria
  - Comparability of groups through randomization
  - Clearly defined treatment strategy
  - Clearly defined clinical outcomes (methods, timing, etc.)
  - Unbiased ascertainment of outcomes (blinding)
  - **Prespecified primary analysis**
    - Sampling plan
    - Population analyzed as randomized
    - Summary measure of distribution (mean, proportion, etc.)
    - Adjustment for covariates

42

Why Confirmatory RCT?

- Multiple RCT contradicting observational studies in real life
  - CAST, CARET, WHI
- Exploration of early phase RCT results inflates type 1 error
  - Attempts to address high toxicity, lack of efficacy
  - Modify eligibility, treatment, outcome
- Selection of promising results favors random highs
  - There will be regression to the mean in next study
    - (This of recent interest in group sequential tests)
  - Inclusion of preliminary result will bias any "meta-analysis"

43

U.S. Regulation of Medical Devices

- Medical Devices Regulation Act of 1976
  - Class I: General controls for lowest risk
  - Class II: Special controls for medium risk - 510(k)
  - Class III: Pre marketing approval (PMA) for highest risk
    - "…**valid scientific evidence** for the purpose of determining the safety or effectiveness of a particular device … adequate to support a determination that there is reasonable assurance that the device is safe and effective for its conditions of use…"
    - "Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, **from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness**…"
- Safe Medical Devices Act of 1990
  - Tightened requirements for Class 3 devices

44

Optimizing the Process

- How do we maximize the number of drugs adopted while
  - Ensuring effectiveness of adopted drugs
  - Ensuring availability of information needed to use drugs wisely
  - Minimizing the use of resources
    - Patient volunteers
    - Sponsor finances
    - Calendar time
- The primary tool at our disposal: Sequential sampling
  - Decrease average sample size used for each drug screened
  - Maximize number of new drugs using limited resources

45

Science and Statistics

- Statistics is about science
  - (Science in the broadest sense of the word)
- Science is about proving things to people
  - (The validity of any proof rests solely on the willingness of the audience to believe it)

46

Phases of Investigation

- A sequential, adaptive process
  - But only “piecewise continuous”
- During any individual clinical trial
  - Sequential monitoring, adaptation addresses that trial’s issues
- “White space” between trials: Detailed and exploratory analyses
  - Evaluation of multiple endpoints; cost/benefit tradeoffs
  - Exploratory analyses
  - Integration of results from other studies
  - Management decisions
  - Regulatory and ethical review
- Next RCT (if any): May address different question or indication

47

Classifying Methods

- Sequential approaches across studies (phases)
  - Status quo: Phase 1, 2, 3 with multiple confirmatory trials
  - Seamless phase 2/3
  - Pivotal studies
  - Accelerated approval
- Sequential approaches within a study
  - Fixed sample
  - Group sequential approaches
  - Other adaptive approaches
    - Statistical: sample size re-estimation; repowering
    - Scientific: modifying eligibility (enrichment), treatment, outcome
- Bayesian versus frequentist

48

Overview of Setting

Elements of RCT Design: Specific Aims

49

Phase of Research

- Phase 1: Safety of human experimentation
- Phase 2: Refinement of indication
- Phase 3: Confirmatory trials for regulatory approval
- Phase 4: Post-marketing safety, comparative effectiveness

50

Phase I Clinical Trials

- Initial safety / dose finding in humans
  - E.g., “first in human” studies
- Goals:
  - Pharmacokinetics / pharmacodynamics
  - Incidence of major adverse effects
  - Decide whether it is ethical to continue testing in humans
- Methods
  - Relatively small number of participants
  - Participants often not true target population
  - Sometimes dose escalation
  - Sometimes no comparison group

51

Phase 2 Clinical Trials

- Preliminary evidence of efficacy
- Goals:
  - Screening for any evidence of treatment efficacy
  - Incidence of major adverse effects
  - Decide if worth studying in larger samples
  - Gain information about best chance to establish efficacy
    - Choose population, treatment, outcomes
- Methods
  - Relatively small number of participants
  - Participants closer to true target population
  - Outcome often a surrogate
  - Sometimes no comparison group (especially in cancer)

52

Phase 3 Confirmatory Trials

- The major goal of a "registrational trial" is to confirm a result observed in some early phase study
  - Gain information for regulatory labeling of treatment strategy
- Rigorous science: Well defined confirmatory studies
  - Eligibility criteria
  - Comparability of groups through randomization
  - Clearly defined treatment strategy
  - Clearly defined clinical outcomes (methods, timing, etc.)
  - Unbiased ascertainment of outcomes (blinding)
  - Prespecified primary analysis
    - Population analyzed as randomized
    - Summary measure of distribution (mean, proportion, etc.)
    - Adjustment for covariates

53

Phase 4 Trials

- Examination of incidence of extremely rare, but extremely serious adverse events
  - E.g., major adverse cardiovascular events with Vioxx vs naproxen: 0.4% vs 0.1%
- Evaluation of treatments in broader use
  - Subjects excluded from registrational RCT
  - Vulnerable populations (pregnancy, pediatrics)
- Comparisons among alternative approved therapies

54

Clinical Estimand

- Safety
  - To conduct clinical trials
  - To start therapy
  - To complete therapy
- Efficacy
  - Biologic activity on surrogate versus clinical benefit
  - Subpopulation versus eventual population
  - Short term versus persistent effect
- Effectiveness
  - Risk / benefit tradeoffs
  - Changing patient behavior / ancillary treatments
  - Vulnerable populations
  - Off label use

55

Comment on "Intent to Treat"

- Intent to Treat (ITT) is most often used to mean a per-randomization analysis
  - Subjects analyzed by assigned group regardless of compliance
- There are many problems with the terminology
  - The subjects accrued to the study may not represent the population of all subjects whom might be an eventual target of treatment
  - Any missing data means that true per randomization analysis cannot be performed
- I find it very useful, however, to consider the idea of evaluating safety and efficacy among subjects for whom treatment is recommended
  - This clinical estimand should be considered at least in part for every treatment
  - However, in some settings we may also want to evaluate the treatment effect and safety in subjects who will chronically use it
    - Tolerators, responders, adherers

56

Overview of Setting

Elements of RCT Design: Scientific Design

57

Treatment Groups

- Single arm
- Two arms
- *K* arms
  - Multiple doses
  - Multiple treatments
- Factorial (partial factorial)

58

Assignment of Treatment

- Observational
  - Subject preference
  - Single arm interventions (including “pre-post” designs)
- Randomized
  - Matched controls
    - Cross-over
    - Paired
  - Parallel groups
    - Individual randomization
    - Cluster randomization
    - Step-wedge design

59

Study Subjects

- Eligibility
  - Inclusion criteria: based on disease, patient population or eventual indication
  - Exclusion criteria: based on experimental setting
- Subpopulations
  - Based on pre-randomization variables
  - Based on post-randomization variables

60

Clinical Endpoints

- Method of measurement
- Timing of measurement(s)
  - Calendar time
  - Study time
  - Event time
- Composite endpoints
  - AND: occurrence of all of several events
  - OR: occurrence of at least one of several events
  - SCORE: severity of based on combination of scores for each event

61

Overview of Setting

Elements of RCT Design: Statistical Design

62

Selection of Design - 1

- Statistical endpoints
  - Primary
    - Single endpoint
    - Co-primary endpoint
    - Composite endpoint
    - Multiple endpoints
      - Alternative endpoints
      - Alternative subgroups
  - Secondary, tertiary
  - Exploratory

63

Selection of Design - 2

- Probability model: parametric, semiparametric, distribution-free
- Summary measure
- Quantification of effect
- Statistical analysis model
- Statistical criteria for evidence
- Specification of desired precision
- Sequential sampling plan
- Accrual plan / sample size
- Planning for handling missing data

64

Evaluation of Design - 1

- Frequentist power curve
  - Type 1 error
  - Power under design alternative(s)
- Sample size distribution
  - Maximal statistical information
  - Minimal statistical information
  - Primary endpoints
  - Secondary endpoints
  - Average statistical information
  - Stopping probabilities

65

Evaluation of Design - 2

- Calendar time requirements (means, prediction intervals)
  - Accrual time distribution
  - Total study time distribution
- Potential inference
  - Frequentist
    - Point estimates: consistent, minimal bias
    - Confidence intervals
    - P values
  - Bayesian (for all relevant priors)
    - Posterior mode, mean, median
    - Credible intervals
    - Posterior probabilities

66

Evaluation of Design - 3

- Stochastic curtailment
  - Predictive probabilities of significance
  - Conditional power

67

Overview of Setting

Constraints and Optimality Criteria

68

Design: Distinctions without Differences

- There is no such thing as a “Bayesian design”
- Every RCT design has a Bayesian interpretation
  - (And each person may have a different such interpretation)
- Every RCT design has a frequentist interpretation
  - (In poorly designed trials, this may not be known exactly)
- In this course I focus on the use of both interpretations

69

Application to Drug Discovery

- We consider a population of candidate drugs
- We use RCT to “diagnose” truly beneficial drugs
- Use both frequentist and Bayesian optimality criteria
  - Sponsor:
    - High probability of adopting a beneficial drug (frequentist power)
  - Regulatory:
    - Low probability of adopting ineffective drug (freq type 1 error)
    - High probability that adopted drugs work (posterior probability)
  - Public Health (frequentist sample space, Bayes criteria)
    - Maximize the number of good drugs adopted
    - Minimize the number of ineffective drugs adopted

70

Major Constraints

- I presume the need to meet regulatory criteria
  - However, different RCT will have different goals
- Multiple comparison issues must be addressed
  - Type 1 error rates must be quantifiable
  - Prespecified analysis plans
  - Reliance on “model checking” is generally not acceptable
- Emphasis will therefore be placed on distribution-free probability models
  - However, the relative optimality of such models will (of necessity) be discussed in the context of parametric models

71