Biost 524: Design of Medical Studies

Lecture 3:
Overview of Clinical Trial Design

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

April 5, 2010

1

Lecture Outline

- Goal: Minimizing Bias and Variability
- Protocol
- Defining the Target Population
  - Disease
  - Patient population
- Defining the Intervention(s)

2

Goal of Clinical Trial Design

Minimizing Bias and Variability

Where am I going?

- Establishing the medical value of a new treatment proceeds through a series of investigations in human volunteers
- We thus want to be able to
  - ensure that we answer the important scientific question and
  - minimize number of patients, calendar time, and cost.

3

Clinical Trials

- Goal:
  - Discovery and adoption of new beneficial treatments or diagnostic methods
- Experimentation in human volunteers to investigate a new treatment, preventive agent, or diagnostic method
  - Safety: Do adverse effects outweigh any benefit?
  - Efficacy: Can treatment beneficially alter disease?
  - Effectiveness: Would adoption of the treatment help population's health?

4

Optimality Criteria

- A good procedure will
  - Minimize "false positives"
    - Any treatment recommended for adoption will have a high probability of being a truly effective therapy
  - Minimize "false negatives"
    - Any truly effective therapy will have a high probability of being recommended for adoption
  - Be highly safe and ethical
    - Minimize the number of patients exposed to inferior treatments while investigations proceed
  - Be efficient
    - Minimize costs (patients, calendar time, money)

5

Common Statistical Approach

- Design an RCT to answer relevant question
  - Treatment, patient population, intervention, comparator, outcome
    - There is an underlying probability of our hypotheses being correct: "Prevalence of effective therapies"
- Fix probability of making wrong decisions
  - Erroneously decide against status quo < 2.5%
  - But: erroneously decide against status quo 2.5%
- Design trial to fix sensitivity of study
  - Power: High probability to detect beneficial treatment

6

PV+ and PV- of RCT

- Relationship to type I error, power, and prevalence of truly effective therapies

$$PVP = \frac{Power \times Prev}{Power \times Prev + (Type\ I\ err) \times (1 - Prev)}$$

$$PVN = \frac{(1 - Type\ I\ err) \times (1 - Prev)}{(1 - Type\ I\ err) \times (1 - Prev) + (1 - Power) \times Prev}$$

7

Common Pitfalls of Studies

- Data driven hypotheses
  - Multiple comparisons
  - Over-fitting of data
- Poor selection of subjects, outcomes
- Noncomparability of treatment groups

8

Issues: Bias

- A biased study is one that will systematically tend to estimate a treatment effect that is not correct
  - across replicated experiments (frequentist bias), or
  - with a large sample size (consistency)
- N.B.: The definition of bias is very much dependent upon what we wish we were estimating
  - How are we going to generalize our results?

9

Sources of Bias

- Attributing an observed difference to a particular treatment
  - Disease
    - Misclassification, overly restrictive
  - Patients
    - Insufficiently or overly restrictive
  - Intervention
    - Administered incorrectly, improper restriction of ancillary treatments
  - Comparator
    - Irrelevant comparator, treatment groups not similar
  - Outcomes
    - Irrelevant outcome, measurements differ by group

10

Confounding Bias

- The treatment groups being compared differ with respect to other important (measured or unmeasured) variables that are predictive of outcome
  - Systematic confounding
    - Process of assigning treatments tends to create groups that are dissimilar
      - Patient or provider preference
      - Time trends in diagnosis, treatment
  - Stochastic (conditional) confounding
    - No systematic trends, but we got unlucky this time

11

Ascertainment Bias

- Assessment of outcomes differs across treatment groups
  - Method of measurement
    - Clinical versus subclinical triggers for assessment
  - Frequency of measurement
    - Adverse events leading to higher surveillance
    - Impact on minima, maxima, time to event
  - Misclassification
    - Accuracy and/or precision of measurement affected by treatment (e.g., tumor growth vs inflammation)

12

Effect Modification Bias

- Treatment effect varies across subgroups
  - Can lead to appearance of confounding if subgroup membership differs across treatment groups
  - Also leads to problems in generalizing effectiveness to eventual treated population

13

Reporting Bias

- Tendency to report results agreeing with preconceived notions
  - Publication bias in literature
  - Selection of historical results to get most favorable outcomes
  - Multiple comparison issues in selecting primary outcomes
  - Multiple comparison issues in selecting summary of outcome distributions
- Increases type I error substantially

14

Statistics and Game Theory

- Multiple comparison issues
  - Type I error for each endpoint
    - In absence of treatment effect, will still decide a benefit exists with probability, say, .025
- Multiple endpoints 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

15

Ex: Level 0.05 per Decision

- Experiment-wise Error Rate
-

Number Compared	Worst Case	Correlation 0.00	0.30	0.50	0.75	0.90
1	.050	.050	.050	.050	.050	.050
2	.100	.098	.095	.090	.081	.070
3	.150	.143	.137	.126	.104	.084
5	.250	.226	.208	.184	.138	.101
10	.500	.401	.353	.284	.193	.127
20	1.000	.642	.540	.420	.258	.154
50	1.000	.923	.806	.624	.353	.193

16

For Each Outcome Define "Tends To"

- In general, the space of all probability distributions is not totally ordered
  - There are an infinite number of ways we can define a tendency toward a "larger" outcome
  - This can be difficult to decide even when we have data on the entire population
    - Ex: Is the highest paid occupation in the US the one with
      - the higher mean?
      - the higher median?
      - the higher maximum?
      - the higher proportion making $1M per year?

17

Statistical Issues

- Need to choose a primary summary measure or multiple comparison issues result
- Example: Type I error with normal data
  - Any single test: 0.050
  - Mean, geometric mean 0.057
  - Mean, Wilcoxon 0.061
  - Mean, geom mean, Wilcoxon 0.066
  - Above plus median 0.085
  - Above plus Pr (Y > 1 sd) 0.127
  - Above plus Pr (Y > 1.645 sd) 0.169

18

Statistical Issues

- Need to choose a primary summary measure or multiple comparison issues result
- Example: Type I error with lognormal data
  - Any single test: 0.050
  - Mean, geometric mean 0.074
  - Mean, Wilcoxon 0.077
  - Mean, geom mean, Wilcoxon 0.082
  - Above plus median 0.107
  - Above plus Pr (Y > 1) 0.152
  - Above plus Pr (Y > 1.645) 0.192

19

Issues: Variability

- Even when unbiased, studies that are conducted with low precision present a problem
  - Decreased power leads to decreased positive predictive value of statistically significant results
  - The same number of patients spread across multiple small studies increases the number of statistically significant studies
    - 10,000 pts in 10 studies: Expect 0.25 false positive
    - 10,000 pts in 400 studies: Expect 10 false positive

20

Statistical Design Issues

- Variability of measurements decreased by
  - Homogeneity of patient population
  - Precise definition of treatment(s)
  - Appropriate choice of clinical, statistical endpoints
  - High precision in measurements
  - Appropriate sampling strategy
- NB: But first and foremost, the RCT must be relevant

21

Protocol

Complete Specification of Study Methods

Where am I going?

- A clinical trial is a scientific experiment
- We thus want to be able to adhere to good scientific practice
  - Prespecification of goals and hypotheses
  - Prespecification of materials and methods
  - Prespecification of measurement of outcomes
  - Prespecification of data analysis methods

22

Purpose of Protocol

- We design an experiment to minimize the bias and variability of our measurement of treatment effect
- The protocol documents the ways in which we will conduct our experiment to achieve that goal

23

Study Protocol

- Clinical trial protocol
  - Formal definition of treatments, endpoints, hypotheses, eligibility, study procedures, etc.
  - Serves as
    - Documentation of rationale for study
    - Documentation of prior knowledge
    - Documentation of experimental method
    - Guide to development of manual of operations
    - Guide to development of Statistical Analysis Plan

24

Documentation - 1

- Purpose: Global objective
- Hypothesis: Specific aims
- Materials:
  - Treatment
    - Background
    - Definition
  - Patients (eligibility criteria)

25

Documentation - 2

- Methods:
  - Recruitment / Randomization
  - Treatment schedule
  - Patient monitoring schedule
  - Evaluation criteria
  - Data management
  - Statistical considerations
  - Trial monitoring procedures
- (Results)
- (Conclusions)

26

Example Outline - 1

- Southwest Oncology Group Protocol Format
  - Schema
  - Objectives
  - Background
  - Drug information
  - Staging criteria
  - Eligibility criteria
  - Randomization plan
  - Treatment plan
  - Toxicities to be monitored and dosage modifications
  - Study calendar

27

Example Outline - 2

- Southwest Oncology Group Protocol Format (cont.)
  - Criteria for evaluation
  - Statistical considerations
  - Discipline review
  - Registration guidelines
  - Data submission schedule
  - Special instructions
  - Ethical and regulatory considerations
  - Bibliography
  - Master forms set
  - Appendix

28

Purpose and Hypotheses

- Global goals
  - Need to keep an eye on what we are truly interested in
- Specific aims
  - The specific scientific hypotheses being addressed by this experiment
    - Target patient population
    - Treatment (and comparison)
    - Measure for treatment outcome

29

Defining the Target Population

Inclusion / Exclusion Criteria

Where am I going?

- Patients are the fundamental “material” of our scientific experiment
- We thus want to be able to
  - have a clear definition of the disease we are targeting,
  - exclude patients for whom the risk of RCT is high and
  - for whom the likelihood of successfully completing the RCT is low.

30

Scientific Basis

- A patient population for whom
  - An improved treatment is desired
  - There is no contraindication to the use of the investigational treatment
  - The investigational treatment might reasonably be expected to work
    - Furthermore: the degree of benefit is expected to be nearly the same for all subgroups of patients that can be identified beforehand

31

Clinical basis

- For clinical utility, the definition of the target population must be based on information commonly available prior to start of treatment
  - Definitions based on diagnostic criteria available only after some delay should be avoided
    - e.g., bacterial culture is often only available 24 hours after start of therapy
  - Definitions based on diagnostic tests that are not routinely available should be avoided
    - genetic profile?
    - clinical utility versus basic science

32

Target Population

- Patient population should generally reflect clinical basis as closely as possible
  - Exception: when it is ethical to conduct a clinical trial to answer a basic science question
- Additional concerns in clinical trial setting
  - Clinical equipoise among choice of all possible treatment assignments
  - Conservatism in using untested treatments
  - Patients' compliance with heightened surveillance in a clinical study

33

Documentation

- Precise definition of target patient population is crucial
  - Scientific:
    - Materials and methods of scientific experiment
  - Clincal:
    - Generalization of safety outcomes
    - Generalization of efficacy outcomes

34

Inclusion / Exclusion Criteria

- Inclusion / exclusion criteria define target population
- Source of patients also of great interest for generalizability
  - Primary care versus tertiary care centers' patient populations
  - Regional differences in possible effect modifiers
    - environmental exposures
    - genetic factors

35

Conceptual Framework

- Population of patients with disease
  - Definition of disease by cause vs signs / symptoms
- Subpopulation with disease targeted by intervention
  - I argue "disease" is really defined by treatment
- Subpopulation eligible for study accrual
  - Restricted due to general clinical trial setting
- Eligible patients from which sampled
  - Restricted due to specific clinical trial (location, time)
- Study sample
  - Restricted due to willingness to participate

36

Ideal

- The study sample should look like a random sample from the subpopulation of all diseased patients who would ultimately be judged suitable for the intervention.
  - Negligible impact of restrictions due to clinical trial procedures
  - Negligible impact of restrictions due to locale of clinical trial
  - High participation rate by eligible patients

37

Safety Considerations

- In conduct of clinical trial may want to exclude some patients
  - Need to consider whether at-risk patients should be exposed to unproven therapy
    - Pregnancy
    - Children
    - Liver, renal, heart disease
    - Elderly

38

Safety Considerations

- Generalizing study results: Efficacy vs effectiveness
  - Treatment may have to be delivered to a population larger than studied
    - Diagnostic procedures after approval may be less rigorous
      - Time requirements: Definition of gram negative sepsis
  - "Diagnostic creep"
    - If some disease has no treatment, then there may be tendency to diagnose a disease that does
      - Gram negative sepsis, non VT/VT cardiac arrest
  - Off-label use

39

Inclusion / Exclusion Criteria

- Inclusion criteria:
  - Definition of ultimate target population
- Exclusion criteria:
  - Exceptions required for clinical trial setting
- Above definitions based on my ideal.
  - In fact, the safety and efficacy of the investigation treatment will only have been established in patients meeting both inclusion and exclusion criteria

40

Inclusion Criteria

- Objective criteria of disease
  - Strive for common clinical definitions
  - Minimize subjective criteria
- Measures of severity of disease that might preclude inclusion in target population
  - mild disease might not be of interest
  - severe disease might not be ethical

41

Inclusion Criteria

- Subgroups of interest
  - E.g., age: adult vs children (though avoid unnecessary restriction)
  - E.g., not candidate for surgery or having failed other treatments
  - E.g., genetic subtype
- Contraindications to treatment
  - Ideally, only if ultimate labeling of treatment would include such contraindications
  - E.g., liver disease, renal disease, diabetes

42

Exclusion Criteria

- Contraindications to treatments in clinical trial setting
  - E.g., safety concerns with new drug that might lead to compliance issues with unproven efficacy
  - E.g., contraindication to comparison treatment
  - E.g., language barriers
- Requirements for evaluation of treatment outcome
  - E.g., lack of measurable disease
  - E.g., inability to make clinic visits
  - E.g., simultaneous participation in other clinical trials

43

Exclusion Criteria

- Requirements for compliance to protocol
  - E.g., not passing a run-in period
  - (but need to avoid lessening generalizability)
- Requirements for ethical investigation
  - unwillingness or inability to provide informed consent

44

Comments re Specification

- Criteria for inclusion / exclusion should consider
  - Methods of measurement
  - Need for and impact of multiple measurements
    - effect of more frequent surveillance
    - possible contradictory measurements
  - Time frames for all criteria
    - usually stated relative to randomization

45

Defining the Intervention(s)

Complete Definition of the Intervention

Where am I going?

- The RCT will ultimately compare outcomes across populations receiving different treatments
- We thus need a prespecification of the interventions, including
  - the nominal intervention,
  - dose modifications, and
  - ancillary prophylactic or rescue treatments.

46

Treatment Strategies

- In human experimentation, we never test a treatment
  - We may not ethically force people to continue a therapy
  - It may not be medically advisable to even want a patient to continue
    - Patients may discontinue a therapy due to headache
    - If forced to continue, those patients may have CVA
- Instead we test a treatment strategy
  - We prescribe an initial treatment
  - Patients may also receive ancillary treatments
    - These may be precipitated by experimental therapy
  - Patients may progress to other therapies

47

Definition of Treatments

- Full description
  - Formulation of treatment
  - Dose, administration, frequency, duration
    - Rules for responsive dosing (e.g., insulin)
    - Include plans for
      - Treatment of adverse events
      - Dose reduction
      - Dose discontinuation
  - Ancillary treatments
    - Prescribed vs allowed vs prohibited
      - (Distinguish safety issues from efficacy issues)

48

Special Issues

- Ultimately, the scientific credibility of the clinical trial stems from our ability to assign a treatment to the participants
  - Ideally we do this in a random fashion
  - At a given point in time, we can only assign a strategy
    - Competing risks may make treatment impossible
    - Intervening events may change indications
    - Informed consent can be withdrawn
  - We must avoid ruining the comparisons of strategies
    - Naïve attempts to compare "treatment" may ruin our ability to assess what really can be tested

49

Ramifications

- Possible actions on progression
  - Stay the course
    - "Progression" dichotomizes a continuous process
    - Treatment may be delaying that process
  - Advance to other therapies
    - Ideally the same for both treatment arms
  - Cross-over to other arm
    - Sometimes motivated to increase sample treated
    - A huge scientific mistake but
      - Ethics sometimes demands it
        - » PA catheterization vs central line
        - » Pemetrexed vs docetaxel

50

Can There Be Noncompliance?

- Experimentally: NO
  - By definition, all patients are following intent to treat
    - Clearly addresses effectiveness questions
    - If efficacy had been our goal:
      - Exclude noncompliant patients as much as possible
      - Increase sample size to deal with attenuation
- Safety: MAYBE
  - We do have to worry that adherence to treatment strategy may change after reporting efficacy
    - We will only have tested safety under the compliance actually achieved
  - Measuring compliance is important for interpretation

51

Ramifications

- An important distinction needs to be made between
  - "Stopping study drug"
    - This may happen due to
      - Adverse events
      - Progression
      - Study burden
    - While we hope for high compliance
      - Badgering patients to remain on therapy can lead to worse adverse events or the quitting the study
    - In the event of stopping study drug, all follow-up of primary outcomes should proceed as planned
  - "Withdrawing consent"
    - No further data will be available

52

Missing Data

- Ideal:

"Just say no."

(Nancy Reagan)

- Real life:

"Missing data happens"

(Bumper Sticker-
rough translation)

53

Types of Missing Data

- Ignorable
  - We can safely throw out the cases with missing data without biasing our results
- Nonignorable
  - Omitting cases with missing data leads to erroneous conclusions

54

Solutions?

"If certain girls don't look at you
It means that they like you a lot
If other girls don't look at you
It just means they're ignoring you
How can you know, how can you know?
Which is which, who's doing what?
I guess that you can ask 'em
Which one are you baby?
Do you like me or are you ignoring me?"

Dan Bern, *"Tiger Woods"*

55

Sad Facts of Life

"Bloodsuckers hide beneath my bed"

"Eyepennies", Mark Linkous (Sparklehorse)

- Typically, nothing in your data can tell you whether missing data is ignorable or nonignorable
  - You just have to deal with what you worry about
  - At the time of study design, plans should be made
    - Sensitivity analyses?
      - Worst case for new treatment, best for control; vice versa
    - Imputation?
    - Ignore?

56