Introduction to Clinical Trials - Day 1

Session 3 - Fundamentals of Trial Design

Presented July 23, 2018

Susanne J. May
Department of Biostatistics
University of Washington

Daniel L. Gillen
Department of Statistics
University of California, Irvine

Goals of Clinical Trial Design

Clinical trials

- Experimentation in human volunteers
- Investigation of a new treatment or 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 effect morbidity in the population?

Goals of Clinical Trial Design

A trial must meet minimum scientific standards

- It must address a meaningful question
  - Discriminate between viable hypotheses (Science)
- Trial results must be credible to the scientific community
  - Valid materials, methods (Science, Statistics)
  - Valid measurement of experimental outcome (Science, Clinical, Statistics)
  - Valid quantification of uncertainty in experimental procedure (Statistics)

Goals of Clinical Trial Design

Individual Ethics

- Conducted in human volunteers, the clinical trial must be ethical for participants on the trial
  - Minimize harm and maximize benefit for participants in clinical trial
  - Avoid giving trial participants a harmful treatment
  - Do not unnecessarily give trial participants a less effective treatment

Goals of Clinical Trial Design

Group Ethics

- The clinical trial must ethically address the needs of the greater population of potential recipients of the treatment
  - Approve new beneficial treatments as rapidly as possible
  - Avoid approving ineffective or (even worse) harmful treatments
  - Do not unnecessarily delay the new treatment discovery process

Goals of Clinical Trial Design

Optimality criteria

- A good procedure will

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

Goals of Clinical Trial Design

Role of Statistics

- Answering scientific questions in presence of variable response
- Scientific questions often reduce to comparing the magnitude of some measurement across groups
- Outcome measures are rarely constant
  - Inherent randomness
  - Hidden (unmeasured) variables
- Use of probability models for describing variability in the real world
  - Distribution of measurements
  - Summary measure (functional) for scientific tendency
  - Quantification of uncertainty in (contrast of) functional(s) (Signal and noise)

Goals of Clinical Trial Design

Common statistical approach

- Optimality criteria (1) and (2) speak directly to the need for achieving high PPV and low NPV
- 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

Positive predictive value in research

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

$$PPV = \frac{\text{Power} \times \text{Prev}}{\text{Power} \times \text{Prev} + (\text{Type I Error}) \times (\text{1-Prev})}$$

$$NPV = \frac{(\text{1-Type I Error}) \times \text{1-Prev}}{(\text{1-Type I Error}) \times \text{1-Prev} + (\text{1-Power}) \times \text{Prev}}$$

Goals of Clinical Trial Design

Predictive value of statistically significant result depends on

1. Probability hypothesis is true to begin with (start with “good ideas")
   - Fixed when hypothesis is formulated
2. Type I error (Specificity)
   - Fixed by level of significance
3. Power (Sensitivity)
   - Statistical power made as high as possible by design

Goals of Clinical Trial Design

The later two elements are improved by

1. Minimizing bias
   - Remove confounding and account for effect modification
2. Decreasing variability of measurements
   - Homogeneity of population, appropriate endpoints, appropriate sampling strategy, more precise measuring device

Goals of Clinical Trial Design

Common pitfalls of studies

- Common pitfalls of experimentation are:
  - Data driven hypotheses (↑ Type I error)
  - Multiple comparisons (↑ Type I error)
  - Poor selection of subjects (↓ Power)
  - Over-fitting of data (↑ Type I error, (↓ Power)
  - Poor selection of subjects, outcomes (↓ Power)
  - Noncomparability of treatment groups (↑ Type I error)
- Each of these pitfalls leads to increases in variability and/or bias in clinical trials...

Variability

How does variability arise?

- Intuitively, if the same experiment is performed several times, the observed results will differ each time
- This variability in observed response depends on several factors including:
  1. The homogeneity of trial participants
  2. How consistently treatment is administered
  3. How consistently the response is measured
  4. Sample size

Reducing Variability

Increasing homogeneity of trial participants

- Inclusion/exclusion criteria to identify a population for whom
  - A new treatment is needed
  - Experimental treatment is likely to work
    - Expected to work equally well in all subgroups
  - All patients likely to eventually use the new treatment are represented (safety)
- Ex: A patient which allows only patients with limited disease and a ECOG score of 0-1 will be much less variable thane one which allows any extent of disease and a ECOG score of 0-4

Reducing Variability

Increased reliability of response measurement

- Objective response measurements and consistent reproducible measurements are critical
- Reducing the subjectivity of response assessment (inter- and intra-rater reliability) will decrease variability
- For biomarkers, use of a single assay analyzed as a single laboratory will decrease measurement error

Reducing Variability

Adequate sample size

- Statistical information is heavily dependent up the number of independent sampling units
- A larger number of patients will lead to reduced variability
- The result is a more precise estimate of treatment effect
- Note: Increasing the number of measurements on a given patient does not contribute the same amount of information as increasing the number of independent patients

Bias

What is bias?

- In statistics, bias is a tendency of a statistical estimate to deviate in one direction from a “true value"
  - What defines the “truth" is dictated by the scientific goal
- 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)
- As in the statistical definition, the definition of a biased study is very much dependent upon what we wish we were estimating
  - How are we going to generalize our results?

Sources of Bias

- Attributing an observed difference to a particular treatment
  - Disease
    - Misclassification, overly restrictive
  - Patients
    - Insufficiently selected 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

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

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)

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

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

Multiple Comparisons

Multiple comparisons

- 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

Multiple Comparisons

Multiple comparisons

“When you go looking for something specific, your chances of finding it are very bad, because of all the things in the world, you’re only looking for one of them.

“When you go looking for anything at all, your chances of finding it are very good, because of all the things in the world, you’re sure to find some of them."

- Darryl Zero in “The Zero Effect"

Multiple Comparisons

Multiple comparisons

In Statistics-Speak “When you go looking for something specific, your chances of finding [a spurious association by chance] are very bad, because of all the things in the world, you’re only looking for one of them.

“When you go looking for anything at all, your chances of finding [a spurious association by chance] are very good, because of all the things in the world, you’re sure to find some of them."

Multiple comparisons

- Goal is to achieve reproducible scientific evidence, but multiple comparisons lead to
  - Inflation of type I error rates
  - Spurious associations
- Consider the experiment-wise type I error rate as a function of the number of comparisons and the correlation between endpoints

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

Multiple comparisons

- Some believe that this problem only exists when testing different outcomes
- However, the issue also exists when testing multiple summary measures for the same outcome!
- As an example, consider the type I error for a two group comparison of a normally distributed outcome

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

- Bottom line: Need to specify a primary summary measure or multiple comparison issues result!

Summary Remarks

Essentials of trial design

- A good procedure will

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

Summary Remarks

Essentials of trial design

- First and foremost, the trial must be relevant
  - Conducted in appropriate patient population (new treatment needed and likely to work)
  - Testing appropriate hypothesis
- Predictive value of trial results is increased by
  - Decreasing variability:
    - Homogeneity of patient population
    - Precise definition of treatment(s)
    - Appropriate choice of clinical, statistical endpoints
    - High precision in measurements
    - Appropriate sampling strategy
  - Minimizing bias:
    - Use of appropriate comparison group
    - Blinding
    - Use of randomization
    - Avoiding multiple comparisons

Summary Remarks

We're not alone...

- International Conference on Harmonisation (ICH: www.ich.org):
  - Launched in 1990: a harmonization of requirements for pharmaceutical registration in US, Europe, and Japan.
  - An excellent resource for current best practice.
- ICH Part E9 - Statistical Principles.
- CONSORT guidelines
  - An agreement between major journals on standards of evidence.
    * The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration.
    * Douglas G. Altman, DSc; Kenneth F. Schulz, PhD; David Moher, MSc; Matthias Egger, MD; Frank Davidoff, MD; Diana Elbourne, PhD; Peter C. Gotzsche, MD; and Thomas Lang, MA, for the CONSORT Group.
    * Ann Intern Med. 2001;134:663-694.
    * http://www.consort-statement.org

Defining the Target Population

Inclusion/Exclusion Criteria

- Patients are the fundamental “sampling units" 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 likelihood of successfully completing the RCT is low

Defining the Target Population

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

Defining the Target Population

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

Defining the Target Population

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

Defining the Target Population

Inclusion/exclusion criteria

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

Defining the Target Population

Conceptual framework

- Population of patients with disease
  - Definition of disease by cause vs signs / symptoms
- Subpopulation with disease targeted by intervention
  - 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 (efficacy vs. effectiveness)

Defining the Target Population

Ideal study sample

- 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

Defining the Target Population

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, co-morbidities, elderly
- Generalizing study results: Efficacy vs effectiveness
  - Self-selection into trial
  - Treatment may have to be delivered to a population larger than studied
    - Diagnostic procedures after approval may be less rigorous (eg. time requirements in definition of gram negative sepsis)
  - Off-label use

Defining the Target Population

Inclusion/exclusion criteria

- Inclusion criteria:
  - Definition of ultimate target population
- Exclusion criteria:
  - Exceptions required for clinical trial setting
- The safety and efficacy of the investigation treatment will only have been established in patients meeting both inclusion and exclusion criteria

Defining the Target Population

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

Defining the Target Population

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

Defining the Target Population

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
- 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

Defining the Target Population

Specification of inclusion/exclusion criteria

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

Defining the Intervention

Treatments vs. treatment strategies

- The trial will ultimately compare outcomes across populations receiving different treatments
- In a clinical trial, 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
- 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

Defining the Intervention

Definition of treatments

As such, a full description of the treatment is necessary

- 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)

Choice of Outcome

Clinical outcomes

- Goal of a clinical trial is to establish if an experimental treatment will prevent a particular clinical outcome:
  - Development of disease
  - Decreased quality of life
  - Mortality
- Essential to define relevant outcome and summary measure
  - Probability of mortality within 28 days
  - Number of days alive and out of ICU
  - Mean 6-minute walk distance
  - Median survival

Choice of Outcome

Clinical outcomes

- A common problem is that the clinical outcomes are rare or occur after a long time.
- This has an impact on trial design:
  - Larger sample sizes are required to detect treatment effects on rare events.
  - Long periods of follow-up may be needed to assess clinical endpoints.

Choice of Outcome

Surrogate outcomes

- A surrogate outcome is a biological endpoint which:
  - Can be measured in a shorter time frame
  - Can be measured precisely
  - Is predictive of the clinical outcome.
- Use of a surrogate may increase trial efficiency.
  - Assume that treatment effect on the surrogate is a good indication of its effect on the clinical outcome

Choice of Outcome

Examples of surrogate outcomes

- AIDS:
  - HIV leads to depression of CD4 cells
  - Increased viral load correlates with development of AIDS
  - Surrogate endpoint: viral load
  - Clinical endpoint: morbidity and/or mortality

- Coronary heart disease:
  - People with arrhythmia following heart attack (MI) have poor survival.
  - Therapies have been developed toward preventing arrhythmia.
  - Surrogate endpoint: arrhythmia
  - Clinical endpoint: mortality.

Choice of Outcome

Clinical endpoints

- Clinical versus biological (surrogate) endpoints
  - Typically, subjects participating in a trial are hoping that they will benefit in some way from the trial
  - Clinical endpoints are therefore of more interest than purely biological endpoints
  - For late stage trials, how well does the proposed surrogate correlate with the targeted clinical endpoint?
  - Often there is great potential for being led astray by a surrogate outcome which may pose safety issues
- More later (and Day 2!)

Comparison Groups

Need for Comparison Groups

- Clinical trials can utilize:
  - No comparison group
  - Historical controls
  - Concurrent comparison group(s)
- Having a comparison groups is important when
  - Deciding whether a proposed treatment is effective
  - Deciding among the alternatives when treating a single patient

Comparison Groups

No comparison group

- Appropriate when an absolute criterion for treatment exists
- Single arm clinical trial
  - Cohort design
  - Includes “pre-post" designs
- Rarely do such absolute criterion exist. Instead, we are really invoking the the results from previous investigations
  - Ex: Pearl Index of 2.0 for evaluation of oral contraceptives

Comparison Groups

Historical controls

- An attempt to make more efficient use of limited research resources
- Single arm clinical trial
- Compare results to
  - Absolute criterion derived from historical trials
    - Dishonest : Treat historical estimates as known and use only one-fourth the sample size compared to a 2-arm study
  - Sample from historical clinical trial (better)
    - More honest : Account for variability in historical control estimate save only half the sample size relative to a 2-arm study

Sample size requirements

- Consider generic sample size formulae for the no control, historical control, and concurrent comparison groups
- Sample size requirements in a single arm study to detect a mean outcome greater than μ_0

$$n = \frac{(z_{1-\alpha/2} + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

- Sample size requirements on experimental arm in a two arm study to detect a mean outcome greater than μ_0

$$n_1 = \frac{(z_{1-\alpha/2} + z_\beta)^2 \left(1 + \frac{n_1}{n_0}\right) \sigma^2}{(\mu_1 - \mu_0)^2}$$

Sample size requirements

- Sample size requirements on experimental arm when using historical controls in a study to detect a mean outcome greater than μ_0

$$n_1 = \frac{(z_{1-\alpha/2} + z_\beta)^2 \left(1 + \frac{n_1}{n_0}\right) \sigma^2}{(\mu_1 - \mu_0)^2}$$

  - n_0 historical controls are presumably already available

Comparison Groups

Use of historical controls

- Thus, compared to a two arm study of a new treatment and a historical treatment, use of historical controls can save time and money
  - Use of historical control sample obviates the need for one arm; thus only half the subjects when 1:1 randomization utilized
  - Using the estimates from a historical clinical trial as if they were known treatment effects decreases sample size requirements even further:
    - Only one-fourth the number of subjects are required
    - However, we are pretending that we have an infinite number of relevant historical controls (no variability)!

Use of historical controls

- However, the validity of such methods is heavily dependent upon the historical trial being comparable in every way
  - No changes in comparison treatment
  - No changes in definition of study population
  - No changes in ancillary treatments
  - No changes in measurement of treatment outcome
- Pocock (J Chronic Disease, 1976) described conditions for acceptability of historical control group

Comparison Groups

Pocock conditions for use of historical controls

1. Such a group must have received a precisely defined standard treatment
   - relevance of standard treatment must remain
   - measurement of treatment parameters must be the same
   - ancillary treatments must not have changed
2. Group must have been a part of a recent clinical study containing the same requirements for patient eligibility
   - measurement methods used in eligibility must be the same
   - clinical trial setting must have same selection pressures on patient participation
3. Methods of treatment evaluation must be the same
   - same criteria (schedule) for performing evaluations
   - same criteria for judging outcomes

Comparison Groups

Pocock conditions for use of historical controls (cont'd)

4. Distributions of important patient characteristics should be comparable
   - same univariate distributions of risk factors (within range dictated by eligibility criteria)
   - same correlations among risk factors
   - must hold for both measured/unmeasured risk factors of
     - disease, adverse outcomes, and competing risks
5. Previous study must have been performed in the same organization with largely the same clinical investigators
   - must control any subjective aspects of definition of eligibility, treatments, outcome
   - must control for unique patient populations due to location and/or referral patterns
6. There must be no other indications leading one to expect differing results

Additional criteria for use of historical controls

- The analysis should reflect the variability in the original data, not just the estimates of treatment effect
  - It is “cheating" to pretend there was no variability in assessing the outcome from the historical comparison group.
  - Ideally: use the exact distribution of the covariates
    - Nonlinearities of effects of covariates on outcome and interactions among the covariates might alter the inference

Comparison Groups

Statistical remedies for meeting these criteria?

- Attempts to circumvent some of these requirements using statistical methods
  - Clearly, the above conditions are rarely, if ever, satisfied.
  - Attempts have been made to use statistical models to adjust for differences between the historical control group and a current treatment group.
    1. Adjustment for covariates
    2. Propensity score analysis

Comparison Groups

Adjustment for covariates

- Analysis with adjustment for confounding due to dissimilarities between treatment groups
  - Adjust for important predictors of treatment outcome
  - E.g., analyze treatment effect in a regression model including indicator of treatment
  - include as covariates those prognostic variables that differ between the groups

Comparison Groups

Propensity score analyses

- Propensity score analyses attempt to mimic randomization; does not worry about prognostic capability for outcome
  - Confounding = association between covariate and treatment AND association between covariate and outcome
- Creates a “propensity score" measuring the propensity for an individual with specific covariates to be in the new treatment group
- Perform an analysis adjusting for propensity scores
  - In each stratum, there is no association between covariate and treatment

Comparison Groups

Statistical remedies for meeting these criteria?

- Both approaches suffer from drawbacks noted by Byar (Biometrics, 1980) and Simon (Ca Treat Rep, 1982):
  - The variables that are measured and properly recorded typically explain only a small percentage in the variability in treatment group membership and treatment outcome.
    - That is, the regression models used have a very low R^2, thus our ability to have properly matched groups is rather low.
- Furthermore, progress in diagnostic methods and therapeutic strategies means that few measurements made in the past are exactly comparable to those made now
  - Laboratory and imaging techniques lead to improved diagnosis and staging of disease
    - E.g., earlier diagnosis of disease
    - E.g., detection of metastases at earlier stages causes trends toward milder disease being diagnosed as Stage IV
  - Supportive measures may improve outcomes

Internal controls

- Each subject serves as his/her own control
  - Different treatments at different times
  - Different treatments for different parts of body (eg. eye diseases, skin diseases)
- Note: This does not include “pre-post" designs looking at the change from baseline in a single arm study
  - These would be uncontrolled experiments

Comparison Groups

Concurrent controls

- Two or more treatment arms
  - Placebo or standard therapy
    - “If it is ethical to use a placebo, it is not ethical not to." -Lloyd Fisher
  - Active treatments
    - Sometimes consider equivalence
  - Multiple levels of same treatment
    - Evidence of dose-response
    - Identification of optimal dose

Blinding

What is blinding and how does it differ from concealed allocation?

- *Blinding (or masking)* is when neither the the study subject (single blind) nor the study investigator (double-blind) have knowledge of the treatment being received or delivered.
- *Concealed allocation* is when the study investigator (personnel) do not know the allocation sequence.

Blinding

What is blinding?

- ICH guidelines (part E9):

  Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence which the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of end-points, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.

Blinding

Types of blinding

- Participant and investigator bias can be (and have been) a major source of bias in RCTs
  - Such bias generally stems from knowledge of the type of treatment a participant is assigned in the trial
- In studies with concurrent comparison groups, blinding of treatment assignment can minimize bias
  - Single blind experiments : Participant is unaware of treatment assignment
  - Double blind experiments : Neither the participant nor treatment provider know treatment assignment
  - Triple blind experiments : Monitoring committee also blinded

Goals of Blinding

Consider the scientific objective

- ICH guidelines (www.ich.org) part E9 Statistical Principles

  "The most important design techniques for avoiding bias in clinical trials are blinding and randomisation, and these should be normal features of most controlled clinical trials intended to be included in a marketing application."

- Similar criteria are required in the CONSORT guidelines.

Goals of Blinding

Blinding can serve to

1. Minimize “placebo effect", wherein a participant being treated does better than one not treated, irrespective of the actual treatment
   - This is distinguished from secular trends in outcome that might occur over time (cohort effects)
   - To detect a placebo effect, one can compare a group that unknowingly received placebo to a group that received nothing

2. Minimize investigator bias in assessing
   - Adverse events
   - Treatment outcomes (consider subjective assessments such as time to hemostasis or time to tumor response)

3. Minimize bias due to missing data
   - Patients with chronic disease where multiple competing trials are ongoing may be less likely to continue in a given study with knowledge that they are receiving placebo

Goals of Blinding

Concealed allocation can serve to

- Prevent selection bias attributable to

  1. the participants
  2. the investigator

- “Allocation concealment seeks to prevent selection bias, protects the assignment sequence before and until allocation, and can always be successfully implemented. In contrast, blinding seeks to prevent ascertainment bias, protects the sequence after allocation, and cannot always be implemented."

  (Schultz, JAMA; 1995; 274(18)1456:1458)

Goals of Blinding

Does this really matter?

- Noseworthy (1994).*Neurology* 1994;44:16-20.
  - All patients examined and response judged by both a blinded and unblinded neurologist.

Table 1. *p* Value* of between-treatment comparison of proportion of subjects improved, stable, or worse

Assessment (no. pts)	IV cyclo versus placebo		PLEX versus placebo	
	Blinded	Unblinded	Blinded	Unblinded
6 Months (165)	0.159	0.069	0.246	0.047
12 Months (144)	0.295	0.084	0.086	0.004
18 Months (108)	0.418	0.255	0.106	0.072
24 Months (91)	0.088†	0.152	0.201	0.031
Final (mean, 30.4 months; 165)	0.290†	0.490	0.990	0.590

* Derived from chi-square test of the 2 (treatment) × 3 (improved, stable, worse) frequency table at each assessment point.
† Trend favoring placebo; all other comparisons favor active therapy.
IV cyclo Intravenous cyclophosphamide group (group I).
PLEX Plasma exchange group (group II).

Goals of Blinding

Does this really matter?

- Wright, Am Heart J (1948) 36:801-815.
- Odd/even day allocation in a trial of anticoagulants in MI gave 589 patients in the active treatment arm and 442 patients in the control arm.
  * 57% (589/1031) assigned to active treatment (95% CI for assignment probability: 0.54 to 0.60).
  * Clearly Biased allocation; cannot rule out differences between treatment groups.

Goals of Blinding

Does this really matter?

- Schulz, JAMA (1995) 273(5):408-412.
  * Meta-analysis of 250 trials from Cochrane pregnancy and childbirth database.
  * In trials with inadequate concealment of treatment allocation, odds ratios for treatment benefit were 41% larger (i.e., 41% better) than in trials with adequate concealment:

Table 2.—Odds Ratios in the Unclearly and Inadequately Concealed Trials Compared With Those in Adequately Concealed Trials*

Level of Allocation Concealment	Ratio of Odds Ratios (95% Confidence Interval)	χ^2 (*df*)	*P*
Adequate	1.00 (referent)		
Unclear	0.67 (0.60-0.75)	57.9 (2)	<.001
Inadequate	0.59 (0.48-0.73)		

*Multiple logistic regression model with the dependent variable being binary outcome measures from each meta-analysis. The independent variables included a binary variable for treatment group (experimental vs control); indicator variables to control for the effects of each of the 250 trials; terms for the "meta-analysis by treatment group" interaction to control for the different summary odds ratios for the treatment effects in the 33 meta-analyses; and the "allocation concealment by treatment" interaction terms displayed in this table to analyze their associations with estimates of treatment effects. Model deviance=434.2; *df*=215.

Issues With Blinding

Blinding is not always possible

- Placebo not always possible to be identical in appearance
  - Weight of fiber, viscosity of fluid for injections
- Side effects of treatment may be noticeable
  - Skin discoloration with beta-carotene
  - Injection site reactions
- Burden of treatment may not be ethical
  - Surgery, hospitalizations, repeated radiation exposure from CT scans

Issues With Blinding

Other issue need to be considered

- Appearance of treatments
- Dosage, administration schedules
- Coding and dispensing of treatments
- When and how to unblind
  - Emergent situations
  - Only unblind when treatment of toxicities differs between therapies
- Assessing how well the blind was maintained

Issues With Blinding

Blinded evaluation

- When blinding of participants is not possible, blinded evaluation may be
- Must still ensure a similar schedule of assessments
  - Side effects might lead to more frequent monitoring
- Competing risks (eg. death from other causes) still a problem

Issues With Blinding

Blinded evaluation

- In some cases, use of a blinded independent review committee may be mandated
  - Ex: Progression of disease in the setting of follicular non-Hodgkins lymphoma
  - Investigators at each perform measurable lesion assessments based on CT scans and physical examination to determine response and progression
  - Blinded independent radiology review committee retrospectively read and interpret all CT scans for response evaluation and progression
  - Primary response based upon independent review committee
- Bias and monitoring issues can still arise (cf. Dodd et al, JCO (2008), Brummel and Gillen, OJS (2013))

Blinding

When is blinding unnecessary?

- Blinding is less of an issue with harder endpoints (eg. survival)
- The more objective the measurement of outcome the less important blinding is to the scientific credibility of a clinical trial

Blinding

Subjective outcomes

- In cases where blinding is not possible it is important to make outcome assessments as objective as possible
  - Ex: Hemostatic agents for cessation of minor to moderate bleeding during surgery
  - Control: Sponge; Treatment: Powder
  - Not possible to blind surgeon
  - Surgeon responsible for determining when hemostasis has occurred
  - How to define hemostasis?

Treatment allocation

Objective, need, requirements

- Objectives:
  - Treatment groups must be comparable so that differences between groups are due to treatment.
  - Assure against confounding (by both measurable and unmeasurable differences):
    - We might be able to adjust for confounders that can be measured.
    - We cannot adjust for unmeasured differences.
    - To measure confounders we would have to know them *a priori*.
- Requirement:
  - Randomization assures that on average all treatment groups are comparable.

Treatment allocation

Methods

- Concealed allocation:
  - Study personnel cannot determine the treatment assignment before it occurs (or not until study completion in double blind trials).
  - Bias can occur with inadequate concealment.

Treatment allocation

Methods

- Approaches to Randomization
  - Completely randomized designs
  - Blocked randomization
  - Stratified randomization
  - Baseline-adaptive randomization
  - Response-adaptive randomization
  - Cluster randomization

Treatment allocation

Methods: Completely randomized designs

- Treatment assignment is made by randomly allocating a subject to one of the treatment groups without considering previous treatment allocations or the subject's covariates.
  - With equal probabilities of getting any one of the treatments (like flipping a coin).
  - With unequal probabilities of getting each of the treatments (like flipping a biased coin).
- Advantages:
  - Analysis is straightforward
  - Simple to implement
- Disadvantages:
  - In small trials this may result in loss of power and/or bias due to:
    - Unequal number of subjects on treatment arms.
    - Imbalances in the types of patients on different arms.
    - Time trends in non-study treatments or types of patients.

Methods: Blocked Randomization

- Random treatment allocation in (relatively small) blocks so that the desired number of subjects in each treatment is assured.
  - E.g., If you want 500 subjects in each of two treatment groups, then assign patients in 50 blocks of 20 patients so that in each block 10/20 are assigned to each treatment
- Advantages:
  - Potential for more power due to equal number of patients on each arm.
  - Better protection against time trends.
- Disadvantages (none really, but...):
  - (Analysis could account for blocking to attain higher power.)
  - (More complicated to implement and rarely done.)

Treatment allocation

Methods: Stratified Randomization

- Randomization in strata defined by important covariates:
  - E.g., To guarantee gender balance, randomize in small blocks in males and females separately (e.g., first 20 males are equally allocated between treatments A and B; first 20 females are equally allocated between treatments A and B).
  - Particularly useful in small trials with a few covariates that are strong predictors of outcome.
  - Difficult with small numbers in each strata (i.e., cannot have a large number of stratification variables).
- Advantages:
  - Guarantees balance on important covariates (reduces chance of confounding).
  - Reduces variation
- Disadvantages:
  - More difficult to implement
  - Analysis should account for stratification variables (adjust for stratification variables).

Treatment allocation

Methods: Baseline-Adaptive Randomization

- Adaptively modify the randomization procedure to ensure comparable frequency distributions of several covariates.
  - E.g., if there is currently an excess of males receiving treatment A, then the next male should be assigned to treatment B.
  - (Minimization) Each patient is allocated to minimize the imbalance between all important covariates.
- Advantages:
  - Same as for stratification
  - May work better in small samples
- Disadvantages:
  - Much more difficult to implement.
  - Analysis must account for the covariates that controlled the allocation.

Methods: Response-Adaptive Randomization

- “Play the winner" designs:
  - Modify the number of subjects assigned to each treatment according to outcomes of previous subjects.
  - You must have knowledge of previous outcomes by treatment group for each randomization.
- Advantages:
  - Decreases the number of subjects who receive an inferior treatment.
- Disadvantages:
  - May decrease power of the study (serious imbalances may result).
  - Increased chance for bias.
  - May not convince the scientific community.

Treatment allocation

Response-Adaptive Randomization (Example)

ECMO: Extracorporeal Membrane Oxygenation in neo-natal respiratory failure.

- Persistent pulmonary hypertension results in right to left shunt through the foramen ovale or ductus arteriosus causing hypoxemia. ECMO is used to maintain life until the condition resolves.
- Trial 1 (Play the winner absolutely): *Pediatrics* (1985) 76:479-487
  - First subject was randomized to conventional medical therapy (CMT); the infant died.
  - Second subject given ECMO; infant lived.
  - Next 8 subjects given ECMO; all lived.
  - Result:
    100% mortality with CMT
    0% with ECMO
    RR = 0.

Response-Adaptive Randomization (Example)

ECMO Example (con't):

- Trial 2 (Play the winner with higher probability): *Pediatrics* (1989) 84(6):957-63
  - Randomize until the 4th CMT death, then treat remainder with best approach.
  - 19 babies in first phase (4/10 die with CMT; 0/9 die with ECMO).
  - 20 babies on ECMO in second phase (1 death).
  - Result:
    40% (4/10) mortality with CMT;
    3% (1/29) with ECMO;
    RR = 0.086.

- Trial 3 (conventional RCT): *Pediatrics* (1998) 101(4):E1
  - Randomize 185 infants (92 to CMT, 93 to ECMO)
  - Result:
    59% (54/92) mortality with CMT;
    32% (30/93) with ECMO;
    RR = 0.55.

Response-Adaptive Randomization (Example)

ECMO Example (con't):

Implications of the ECMO example:

- ECMO looked better with response-adaptive randomization.
- Response-adaptive designs were not accepted as adequate justification for ECMO.
- Inadequate study designs can delay introduction of beneficial treatments or prolong use of inferior treatments.

"In fact, in the ECMO trial, the patient who failed on treatment B had the most extreme values on no fewer than four important covariates (Paneth & Wallenstein, 1985), and was clearly the sickest. In effect, the trial provides no information whatsoever regarding the treatment comparison. "

-Begg (1990)

Response-Adaptive Randomization (Example)

- The ECMO experience has tempered enthusiasm for randomized PTW
- This being said, there may be times were response-adaptive randomization will work, but
  - There needs to be a clear dilemma re individual ethics
  - There will tend to be decreased group ethics
  - It takes a lot of planning in order to obtain results that will be sufficiently credible

Treatment allocation

Methods: Cluster Randomization

- Subjects are randomized in groups (e.g., churches, schools, cities).
  - Useful when treatment cannot be administered on an individual level without contamination (e.g., smoking cessation studies).
  - Often clusters are matched and treatments are assigned within the matched pairs.
- Advantages:
  - Allows investigation of community interventions.
  - Eliminates contamination bias.
- Disadvantages:
  - Sample size is the number of clusters not the number of individuals.
  - May lose power over individual randomization.

Treatment allocation

Methods: Logistics of Randomization

(a) Completely randomized designs:
- Create column from 1 to $2N$.
- Create column of random numbers uniformly distributed between 0 and 1.
- If the random number is less than 0.5, then the subject receives active treatment, otherwise they receive placebo.

(b) Blocked randomization: For a block of size k with $k/2$ subjects in each of two groups:
- Create a column of $k/2$ A's and $k/2$ B's.
- Create column of random numbers uniformly distributed between 0 and 1.
- Sort the first column according to the second column.
- Repeat for as many blocks as desired.

(c) Stratified randomization: Repeat for each stratum.

Methods: Logistics of Randomization

- Where to perform randomization:
  - Central randomization:
    - Phone calls to the coordinating center.
    - Sequences can be determined at the start of the study (except with adaptive randomization).
  - Distributed randomization: Computer programs, envelopes, or lists at pharmacies.
- Important principles:
  - Strong quality assurance must be in place to ensure proper randomization.
  - Ensure adequate concealment/blinding.
  - Provide for emergency unblinding.
  - Exact randomization scheme must be known for analysis.