

# Experimental Design

## The Key to Reliable & Reproducible Science

Lecture 3:  
Role of Variables, Randomization &  
Common Study Designs

Navneet R. Hakhu, M.S.

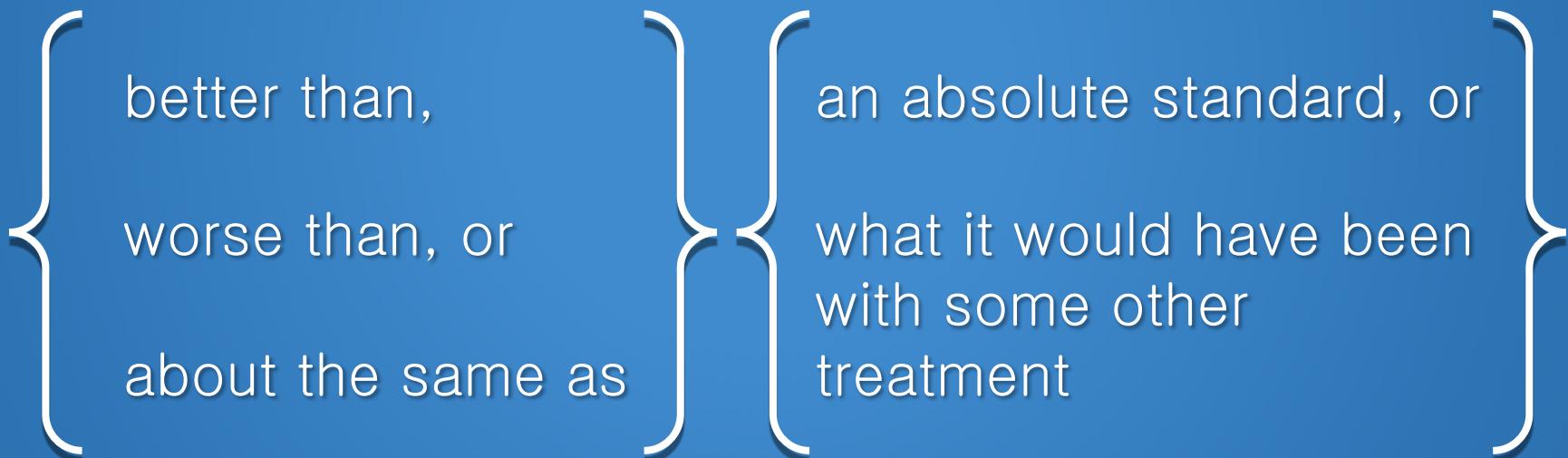
2<sup>nd</sup> Year PhD Student, Statistics

Department of Statistics  
University of California, Irvine

August 28, 2020

# Typical Scientific Hypotheses

- ▶ The treatment will cause an individual's outcome to be



# Counterfactual

- ▶ The statement of the hypotheses assumed that it is possible to know what would have happened under some other treatment
  - Generally we instead have to measure outcomes that are observed
    - In another place (patient)
    - At another time, and / or
    - Under different circumstances

# Causation vs. Association

- ▶ Truly determining causation requires a suitable interventional study (experiment)
  - Comparisons tell us about associations
  - Associations in the presence of an appropriate experimental design allows us to infer causation
    - Be careful: But even then, we need to be circumspect in identifying the true mechanistic cause
      - e.g., a treatment that causes headaches, and therefore aspirin use, may result in lower heart attack rates due entirely to the use of aspirin

# Investigating the Unknown

- ▶ We must acknowledge that we might be wrong
  - It will be impossible to prove something that is not true
  - The treatment might not work as we had hoped

# First Statistical Refinement

- ▶ Determine whether the group that received the treatment will have outcome measurements that are

{ higher than,  
lower than, or  
about the same as } { an absolute standard, or  
measurements in an  
otherwise comparable  
group (that did not  
receive the treatment) }

# Variation in Response

- ▶ There is, of course, usually variation in outcome measurements across repetitions of an experiment
  - Variation can be due to
    - Unmeasured (hidden) variables
      - In the process of scientific investigation, we investigate one “cause” in a setting where others are as yet undiscovered
      - e.g., mix of etiologies, duration of disease, comorbid conditions, genetics when studying new cancer therapies
    - Inherent randomness

## Second Statistical Refinement

- ▶ Determine whether the group that received the treatment will tend to have outcome measurements that are

{ higher than,

lower than, or

about the same as }

} { an absolute standard, or

measurements in an otherwise comparable group (that did not receive the treatment)}

# Evidence Based Medicine

- ▶ Decisions about treatments should consider PICO
  - Patient (population)
  - Intervention
  - Comparators
  - Outcome

# Phase III 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
    - Population analyzed as randomized
    - Summary measure of distribution (mean, proportion, etc.)
    - Adjustment for covariates

# 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: Hormone replacement therapy has lower cardiac morbidity and mortality
  - WHI: Hormone replacement therapy in post menopausal women leads to higher cardiac mortality

# Multiple Comparisons

## ► 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 1 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

# Statistical Role of Variables

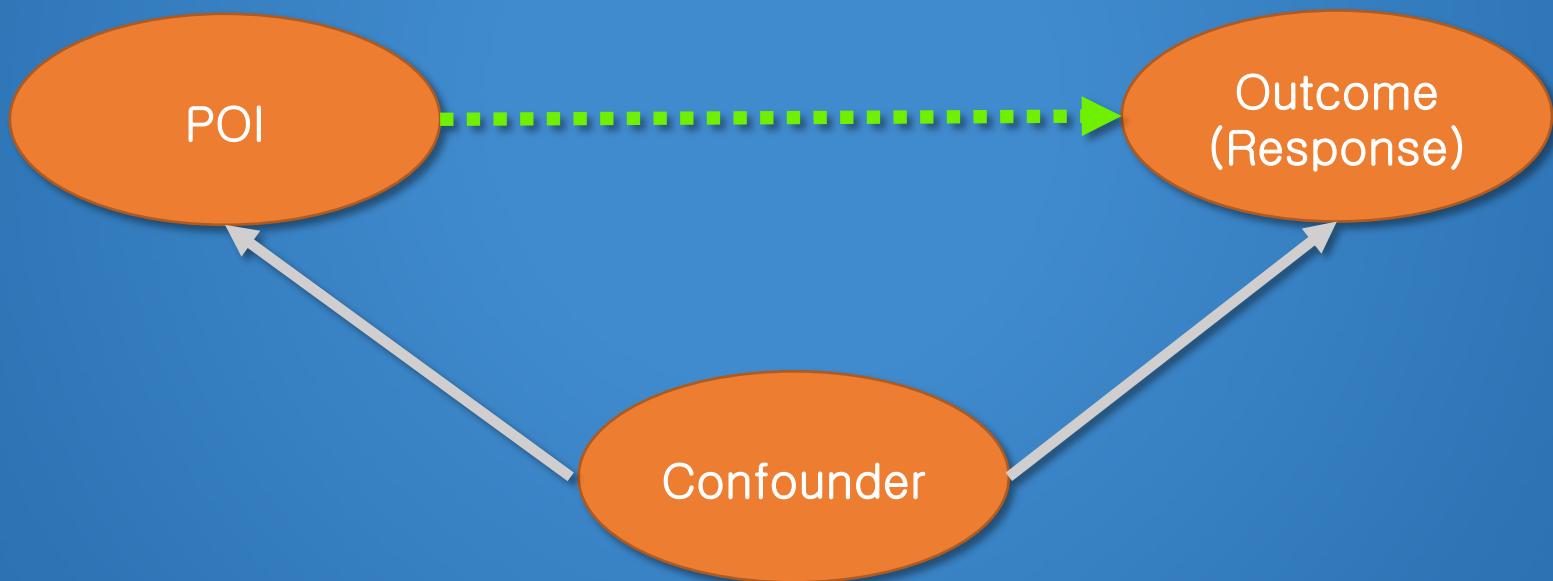
- ▶ Outcome (response) variable(s)
  - Primary and surrogates
- ▶ Predictor(s) of interest (define main groups): POI(s)
- ▶ Subgroups of interest for effect modification
- ▶ Potential confounders
- ▶ Variables that add precision to analysis  
(independent risk factors of response)
  - Known to be associated with response
  - Often these are potential confounders
    - May be associated with predictor(s) of interest in sample
- ▶ Irrelevant to current question

# Confounding

- ▶ Definition of confounding
  - The association between a POI and the response variable is confounded by a third variable if
    - The third variable is associated with the predictor of interest in the sample, AND
    - The third variable is associated with the response
      - Causally (in truth)
      - In groups that are homogeneous with respect to the predictor of interest, and
      - Not in the causal pathway of interest

# Classical Confounder

- ▶ A clear case of confounding is when some third variable is a “cause” of both the POI and response
  - We generally adjust for such a confounder

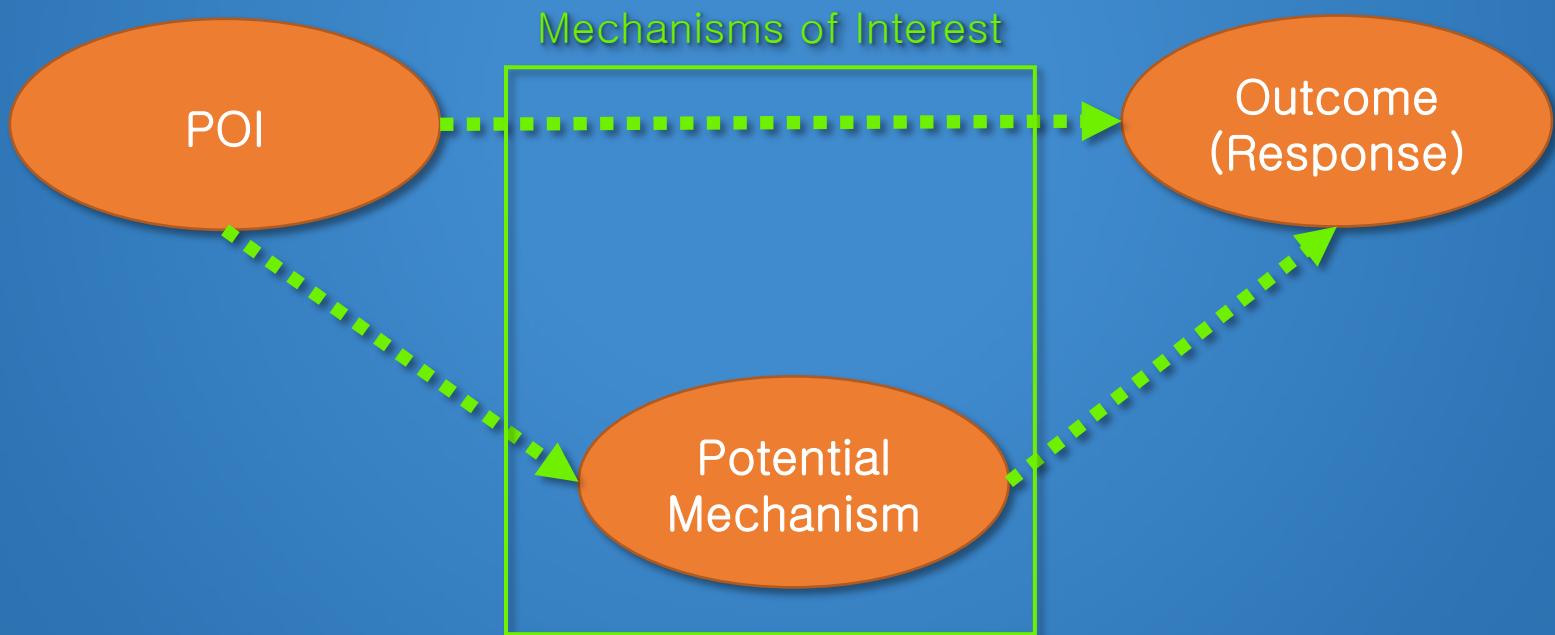


# Adjustment for Covariates

- ▶ We must consider our beliefs about the causal relationships among the measured variables
- ▶ We will not be able to assess causal relationships in our statistical analysis
  - Inference of causation comes only from study design
- ▶ However, consideration of hypothesized causal relationships helps us decide which statistical question to answer

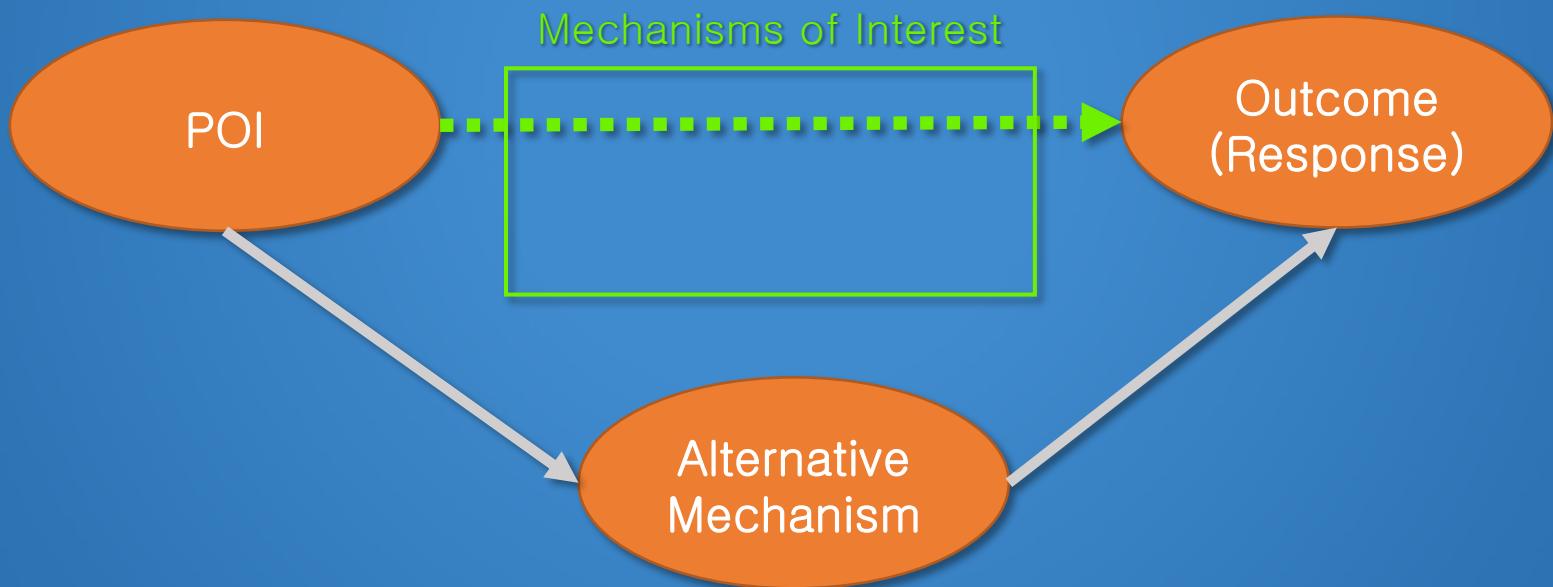
# Causal Pathway

- ▶ A variable in the causal pathway of interest is not a confounder
  - We would not adjust for such a variable (lest we lose ability to detect the effect)



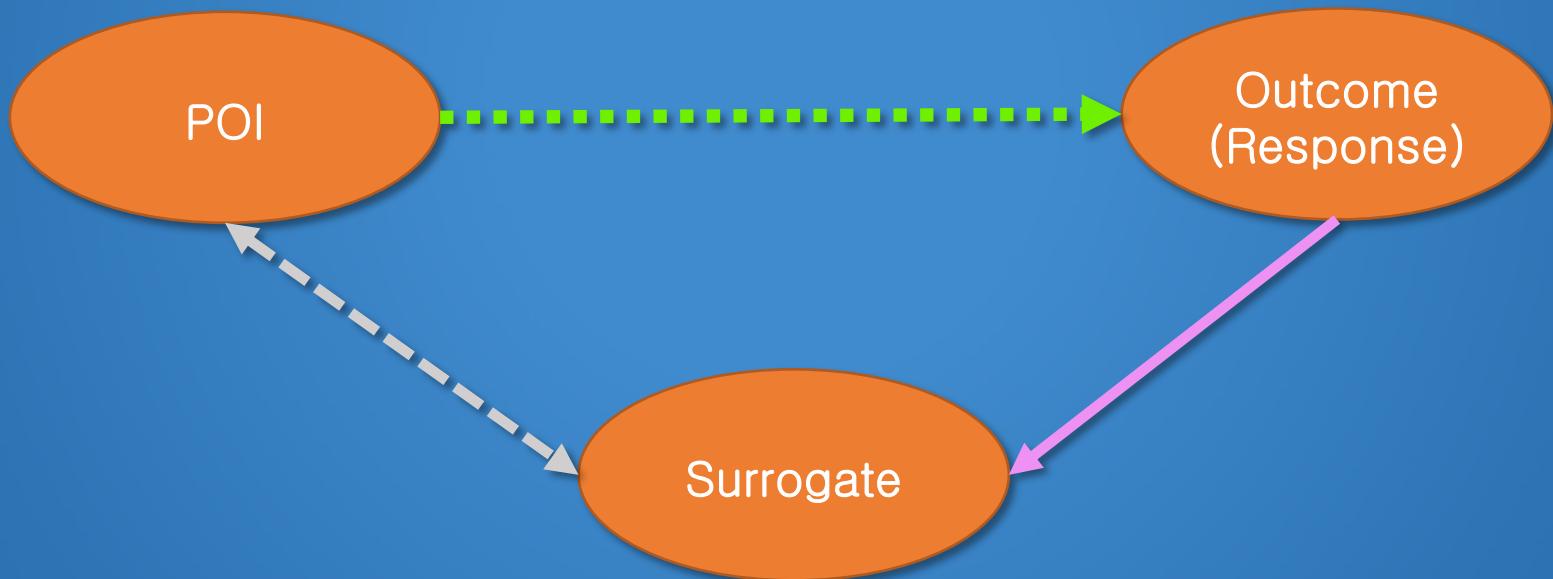
# Causal Pathway

- We would want to adjust for a variable in a causal pathway not of interest
  - e.g., work stress causing ulcers by hormonal effects vs. alcoholism



# Surrogate for Response

- ▶ NOT a confounder: Adjustment for such a variable is a very BAD thing to do



# Unadjusted, Adjusted Analyses

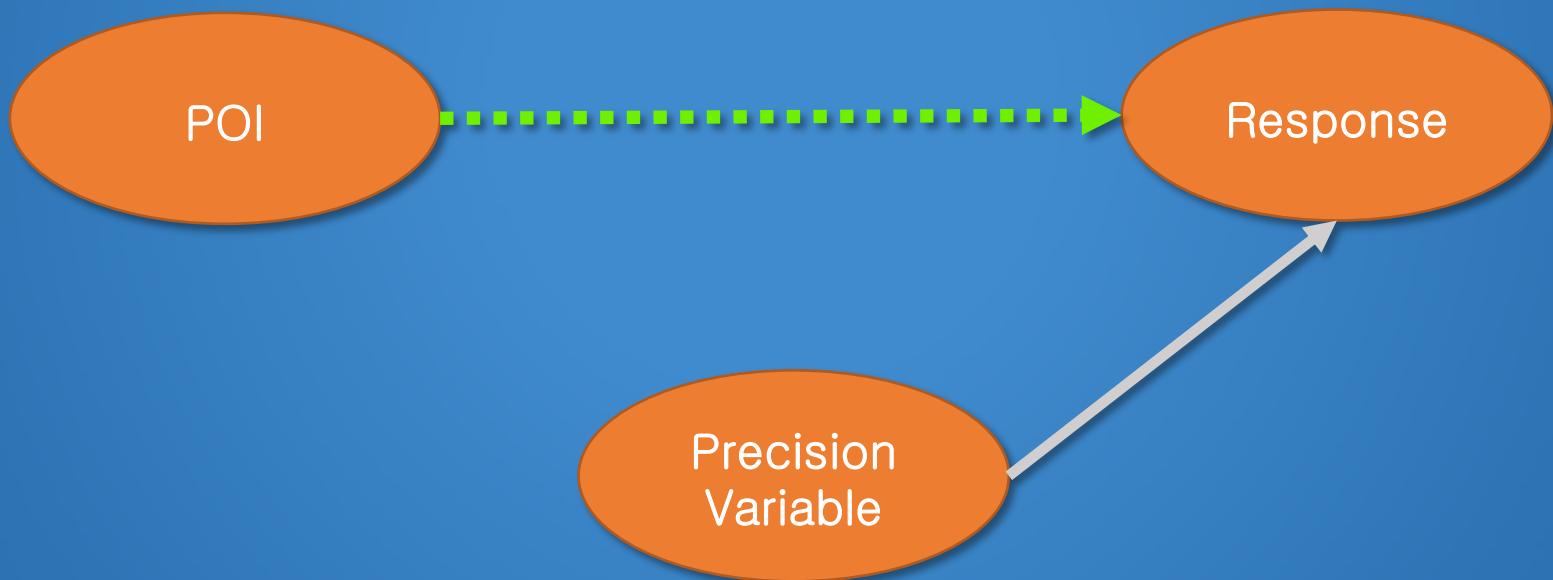
- ▶ Confounding typically produces a difference between unadjusted and adjusted analyses, but those symptoms are not proof of confounding
  - Such a difference can occur when there is no confounding
    - “Precision” variables in logistic, Cox Proportional Hazards regression
    - Complicated causal pathways

# Precision

- ▶ Sometimes we choose the exact scientific question to be answered on the basis of which question can be answered most precisely
  - In general, questions can be answered more precisely if the within group distribution is less variable
    - Comparing groups that are similar with respect to other important risk factors decreases variability
  - Keep in mind: this changes the target population

# Precision Variables

- ▶ The third variable is an independent “cause” of the response
  - We tend to gain precision if we adjust for such a variable



# Controlling Variation

- ▶ In a two sample comparison of means, we might control some variable in order to decrease the within group variability
  - Restrict population sampled (subgroup analysis)
  - Standardize ancillary treatments
  - Standardize measurement procedure

# Effect Modifier

- ▶ The association between Response and POI differs in strata defined by effect modifier
  - Statistical term: “interaction”
- ▶ Depends on the measurement of effect
  - Summary measures
    - Mean, geometric mean, median, proportion, odds, hazard, etc.
  - Comparison across groups
    - Difference, ratio

# Ignoring Effect Modification

- ▶ By design or by mistake, we sometimes do not model effect modification
- ▶ We might perform either
  - Unadjusted analysis
    - POI only
  - Adjusted analysis
    - POI and third variable, but no interaction term
  - Stratified analysis
    - Reweighting to obtain desired average

# Adjusted Analyses

- ▶ If effect modification exists, an analysis only for the third variable (but not interaction) will tend toward a weighted average of the stratum specific effects
  - Hence, an association in one stratum and not the other will make an adjusted analysis look like an association
    - (providing sample size is large enough)
- ▶ Scott Emerson calls this “Intent to Cheat” analysis
  - Gain a larger indication by including strata with no effect

# Analysis of Effect Modification

- ▶ When the scientific question involves effect modification, analyses must be within each stratum separately
  - If we want to estimate degree of effect modification or test for its existence:
    - A regression model will typically include
      - Predictor of interest (main effect)
      - Effect modifying variable (main effect)
      - A covariate modeling the interaction (usually product)
  - If we merely want to ensure treatment effect within each stratum
    - Separate analysis for each stratum
    - (Lower sample size than to prove effect modification)

# Cause and Effect

- ▶ Necessary conditions for establishing cause and effect of a treatment
  - The treatment should precede the effect
    - Beware of protopathic signs
      - News article: Marijuana and risk of MI within 3 hours
        - RCT?
        - Cohort study?
        - Case–crossover design? (Look at case, well were you smoking marijuana beforehand and how much before. Uniformly before that...)
        - Hang–gliding seem to die only when they hang–glide
        - Bias introduced...the earliest sign of a heart attack: ill at ease, some reaction to stress of heart, making you anxious (anxious and smoked or smoked and then had heart attack)
        - Not possible to get through IRB
        - Protopathic – response to feeling ill at ease from heart attack
      - When comparing groups differing in their treatment, the groups should be comparable in every other way

# Major Scientific Tool

- ▶ **Randomization** is the major way in which cause and effect is established
  - Ensures comparability of populations
    - Each treatment group drawn from same population
    - Differences in other prognostic factors will only differ by random sampling
      - Provides balance on the total effect of all other prognostic factors
      - May not provide balance on each individual factor
      - (type 1 error, 5% of some factors will not be balanced)

# Analytic Randomization Models

## ► Randomization model

- Conditions on the sample obtained
  - Only people I care about
  - E.g., permutation tests
  - Pretends that all outcomes were pre-ordained absent a treatment effect
- Tests strong null hypothesis of no treatment effect whatsoever
  - (very unlikely – treatment does something, do benefits outweigh risks?)
  - Under the null hypothesis, any difference in outcome must have been randomization imbalance
    - (too strong an assumption to make)
    - Bone marrow transplantation
      - ...

# Analytic Randomization Models

- ▶ Population model

- Ensures treatment arms drawn from same population initially
- Test weak null hypothesis of no treatment effect on summary measure of interest
  - Can allow for treatment differences between arms on other aspects of outcome distribution

# Analytic Randomization Methods

## ► Comments

- Randomization model does not typically allow testing of non-zero null hypothesis (e.g., non-inferiority)
- Randomization model does not allow distribution-free estimation of confidence intervals
  - For CI, we must know distribution under alternatives
- Assumption of strong null not in keeping with scientific method
  - Assumptions are more detailed than primary question
    - Primary question usually about first moment
    - Semiparametric assumptions are about all moments
  - Consider bone marrow transplantation
- In science we do not try to make more assumptions than we need to make

# Randomization Ratio

- ▶ Most efficient
  - When test statistics involve a sum, choose ratio equal to ratio of standard deviations
  - (most people do not do this)
- ▶ Most ethical for patients on study
  - Assign more patients to best treatment
    - (if we knew what's best, why do a clinical trial?)
    - Many sponsors / patients presume new treatment
    - Adaptive randomization: Play the winner
      - Increase chances of patients on trial to get better treatment (may lead to 10:1 randomization and take longer than 1:1 rand)
      - For general population, PTW may not be good
- ▶ Most ethical for general patient population
  - Whatever is most efficient (generally not adaptive)
  - What is the minimal number of subjects needed to treat (safety data)
- ▶ Other goals
  - Attaining sufficient patients exposed to new treatment
  - Maintaining DSMB blind (if it's not 1:1 rand...)

# Randomization Strategies

- ▶ Complete randomization
- ▶ Blocked randomization (more often)
  - E.g., every 8 patients come in, 4 trt and 4 ctl (protecting against time trends)
  - Ensure balance after every  $k$  patients
  - Ensure closer adherence to randomization ratio
  - Undisclosed block sizes to prevent bias
    - Permutated blocks (random effect like time)
- ▶ Stratified randomization
  - Separately within strata defined by strong risk factors
    - Lessens chance of randomization imbalance
    - By Site (adjust for in the analysis)
  - Need to consider how many variables can be used
    - Cannot have too many because your total sample size may not be large enough
    - Balancing on every combination of covariates (interactions and main effects)

# Points Meriting Repeated Emphasis

- ▶ Clinical trials are primarily concerned with two variables
  - Indicator of treatment
  - Measurement of outcome
- ▶ Additional variables might be of interest
  - Effect modification
    - Treatment effect differs
  - Confounding
    - Variable associated with outcome and treatment
  - Precision
    - Less variability of response within strata
  - (true of any analysis)

# Points Meriting Repeated Emphasis

- ▶ Randomization is our friend...
  - If we randomize, we do not (on average) need to worry about differences between the treatment groups with respect to factors present at time of randomization
    - Any difference in outcomes can be attributed to treatment
      - Again, recognize that treatment can lead to differential use of other ancillary treatments, however
- ▶ But like all friends, we must treat it with respect
  - We must analyze our data in groups defined at the time of randomization
    - Discarding or missing data on randomized subjects may lead to bias
      - It certainly leads to diminished scientific credibility
  - Patients dropping off due to intervention
    - Even if equal numbers does not mean that the reasons are the same
      - Removing subjects can lead to BIAS

# Impact on Data Analysis

- ▶ In presence of randomized treatment assignment
  - Intent to treat analysis (ITT)
    - Based on randomization
    - “Modified ITT” acceptable for efficacy?
      - Efficacy within strata identified pre-randomization
        - Does not violate the randomization
        - Anything after randomization is violated
      - Safety in all subjects
    - Science: Population model (not randomization model)
    - Confounding not an issue (on average)
      - P-value measures probability of observed effects occurring due only to randomization imbalance
    - Gain precision only if adjust for stratification variables
      - Pre-specify that adjustment will be done and HOW it will be done
    - If effect modification is concern --> subgroup analyses

# We want to assess cause and effect

- ▶ How can we do that?
  - One solution: Randomization
    - Why?
    - How?

# Study Designs

- ▶ Observational vs. Interventional (Randomized)
  - Randomization
    - Trying to reduce bias, cause-and-effect, remove confounding on average

# Randomization

- ▶ How to implement?
  - Simple (Complete)
  - (Complete) Blocked
    - Permuted Blocks (Blinding)
    - Incomplete Blocks
  - Stratified
    - Account in the analysis as well
  - Fixed (Non-adaptive) vs. Adaptive

# Study Designs

- ▶ Single Factor (one POI: two levels/groups)
  - Simple Randomized (1:1, r:1)
  - Block Randomized
    - Complete/Incomplete Permuted Blocks
  - Stratified Randomization
  - Paired design (at same time: ophthalmology, psoriasis)
    - Within subject difference
  - Cross-over
  - Repeated Measures
    - Split-Plot
  - Cluster Randomization

# Types of Randomization

- ▶ Simple
- ▶ Block
- ▶ Stratified

# Constrained Randomization: Blocking

## ► Blocking

- To increase precision in any experiment
  - One way
    - Identify and control factor as that contribute to variation in the experimental error
      - Remove source of variation (e.g. sex) by restricting analysis to only females or only males
      - (to reduce heterogeneity, have homogenous group)
      - Disadvantage: sample size reduced (reduces precision); limits generalizability since males have not been considered
  - Divide heterogeneous experimental units into homogeneous subgroups (blocks)
    - Conduct analysis separately within each block
    - Then combine the estimated effects from each block
  - Restrict randomization to within each block separately

# Blinding

- ▶ Motivation for Permuted Blocks
  - Avoiding bias

# Cluster Randomization

- ▶ When treatment cannot be administered on an individual level without contamination
  - e.g., smoking cessation programs
  - e.g., education strategies
  - e.g., out of hospital emergency response
- ▶ Subjects randomized to treatment or control in clusters (the experimental units)
  - Often form matched sets of clusters to randomize in strata

# Cluster Randomization

## ► Advantages

- Allows investigation of community interventions
- Intervention at clinic or village level may be perceived as more ethical
- Logistical considerations for equipment, etc.

## ► Disadvantages

- Sample size may be the number of clusters rather than the number of subjects
- May lose substantial power over randomization by individual

# Cross-over Trials

- ▶ Each subject receives every treatment
  - May gain precision because each subject serves as own control
  - Order of treatments should be randomized
    - A pre-post design is not correctly termed a cross-over design
  - Washout period to avoid carryover effects
    - Analyses should look for differences in treatment effect by order of administration
  - Not feasible with most time-to-event studies

# Cross-over Trials

## ► Advantages

- Greater statistical power in presence of high ratio of between subject to within subject variability in response
  - i.e., when high correlation between repeat measurements of response

## ► Disadvantages

- Cannot be used in presence of
  - Curative treatments
  - Long carryover (and statistical power to detect carryover is usually low)

# Study Designs

- ▶ Multi-factor (2+ POIs: 2+ levels)
  - Factorial design
    - Full vs. Partial
      - Pros, cons

# Factorial Designs

- ▶ Test two or more treatments simultaneously
  - Every subject gets either active or control for each treatment
  - Example: Two treatments: A vs. PlcA and B vs. PlcB
    - Four treatment groups
      - A and B; A and PlcB; PlcA and B; PlcA and PlcB
- ▶ Partial Factorial
  - Some subjects might only participate in one part of the trial
    - Additional treatment groups
      - A only; PlcA only; B only; PlcB only

# Factorial Design Examples

- ▶ Phase II RCT of calcium, fiber for colon cancer prevention
  - Truly looking at biochemical markers of cell proliferation
- ▶ Physicians' Health Study
  - Aspirin (for MI) and beta-carotene (for cancer)
- ▶ Women's Health Initiative (WHI) [partial factorial]
  - HRT (for CVD, breast cancer, osteoporosis, dementia)
  - Low fat diet (for CVD, colon cancer)
  - Calcium / Vitamin D (for osteoporosis, colon cancer)
- ▶ ROC PRIMED in OOH cardiac arrest [partial factorial]
  - Impedance threshold device (individually randomized)
  - Analyze Late vs. Early (cluster randomized)

# Factorial Designs: Settings

- ▶ Completely unrelated treatments and diseases
  - Physicians' Health Study: aspirin for CVD, beta-carotene for cancer
  - Efficiency of clinical trial infrastructure
- ▶ Combination of treatments for same disease
  - Calcium and fiber for colon cancer prevention
  - Allows looking for combined effect
  - However, usually low power to detect effect modification

# Factorial Designs

## ► Advantages

- Answer multiple questions with the same study
  - In absence of effect modification, same power as individual studies
  - Ability to address effect modification (but with low power)

## ► Disadvantages

- Exclusion criteria must consider all treatments
- One treatment may affect compliance on all treatments
- AEs from one treatment may affect ascertainment bias on all treatments

# Large Simple Trials

- ▶ Use many subjects and minimize amount of data collected
  - Definition of treatment must be straightforward
  - Definition of outcome must be straightforward
- ▶ Allows looking at smaller increments of benefit
- ▶ Must not sacrifice scientific rigor, however
  - Ability to assess mechanism of action
  - Ability to detect unexpected toxicity