Biost 524: Design of Medical Studies

Lecture 4:
Comparison Groups; Randomization

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

April 7, 2010

1

Lecture Outline

- Comparison Groups
- Blinding
- Why Randomize?
  - Statistical Role of Variables in Analsis
  - Common Statistical Analysis Models
  - Four Important Questions of Regression

2

Comparison Groups

Options

Where am I going?

- Having a comparison group is important when
  - deciding whether a proposed treatment is effective, and
  - deciding among the alternatives when treating a single patient

3

Comment re Single Arm Trials

"There are only two types of researchers:
- those with a lot of enthusiasm and no controls, and
- those with a lot of controls and no enthusiasm."

(unknown)

4

No Comparison Group

- Appropriate when an absolute criterion for treatment effect exists
- Single arm clinical trial
  - Cohort design
  - Includes “pre-post” designs
- (Rarely do such absolute criteria exist. Instead, we are really invoking the use of results from previous investigations.)

5

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: Use only one-fourth the sample size
  - Sample from historical clinical trial (better)
    - More honest: Maybe only save half the sample size

6

Sample Size: Single Arm Study

- Sample size requirements in a single arm study to detect a mean outcome greater than μ_0

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

7

Sample Size: Two Arm Study

- Sample size requirements <u>on experimental arm</u> in a two arm study to detect a mean outcome greater than μ_0
  - $n_1 = r \times n_0$ with r the ratio between sample sizes

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

8

Sample Size: Historical Controls

- Sample size requirements on experimental arm when using historical controls in a study to detect a mean outcome greater than μ_0
  - n_0 historical controls are presumably already available

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

9

Use of Historical Controls

- Compared to a two arm study of a new treatment and a historical treatment, use of historical data can save time and money
  - Use of a historical control sample obviates the need for one arm: thus only half the subjects when 1:1 randomization.
  - 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
    - But pretending that we have an infinite number of relevant historical controls

10

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

11

Pocock Conditions - 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

12

Pocock Conditions - 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

13

Pocock Conditions - 3

- Methods of treatment evaluation must be the same
  - same criteria (schedule) for performing evaluations
  - same criteria for judging outcomes

14

Pocock Conditions - 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 and unmeasured risk factors of
    - disease,
    - (propensity for) adverse outcomes,
    - and competing risks

15

Pocock Conditions - 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

16

Pocock Conditions - 6

- There must be no other indications leading one to expect differing results

17

Additional Criterion

- 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

18

Proposed Remedies

- 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.
    - Adjustment for covariates
    - Propensity score analysis

19

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

20

Propensity Score Analyses

- Propensity score analyses
  - Attempts 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

21

Drawbacks

- 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.

22

Problem with Time Trends

- 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 may lead to perceived better survival
    - E.g., detection of metastases at earlier stages causes trends toward milder disease being diagnosed as Stage IV
  - Supportive measures may improve outcomes

23

Final Comments re Historical Controls

- The use of historical controls from previous clinical trials would thus appear highly problematic
  - The situation is only worse if one tries to use data from cohort studies or other observational data bases
  - Such data bases are well suited for hypothesis generation and feasibility studies, but do not at all provide comparability to a clinical trial setting

24

Final Comments re Historical Controls

- Nevertheless, when strong treatment effects are expected and when little has changed in the disease setting, the use of historical controls may be the only ethical option
  - Sometimes the window of opportunity for a randomized trial is extremely short
    - Early feasibility studies might show such promising results that preclude equipoise
    - (Some authors therefore suggest randomizing at every stage of investigation)

25

Example: ECMO Study

- Randomized clinical trial of extracorporeal membrane oxygenation in newborns
  - Randomized Play The Winner (PTW) design in which randomization ratio changes to favor more "successful" therapy
- Data:
  - First patient on ECMO survived
  - Next patient on control died
  - Next 9 patients on ECMO survived

26

Comments

- Arguments for credibility of results
  - Prior history suggested 90% mortality under standard of care
- Inference (Begg, 1990)
  - P value of 0.001, 0.051, 0.083, 0.28, 0.62?
- This experience has tempered enthusiasm for randomized PTW
  - Interestingly, follow-up studies had 67% survival on conventional therapy

27

Internal Controls

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

28

Concurrent Control Groups

- Two or more treatment arms
  - Placebo or standard therapy
  - Active treatments
    - Sometimes consider equivalence
  - Multiple levels of same treatment
    - Stronger evidence sometimes obtained from dose-response
    - Identifying optimal dose

29

Blinding

Options

Where am I going?

- Particpant and investigator biases can be (and have been) a major source of bias in RCT

30

Definitions

- 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 participant nor provider know treatment assignment
  - Triple blind experiments:
    - Monitoring committee also blinded

31

Goals

- Blinding can serve to
  - Minimize “placebo effect”: A participant being treated does better than one not being treated, irrespective of the actual treatment
    - This should be distinguished from secular trends in outcome that might happen over time
    - To detect a placebo effect, you compare a group that is unknowingly receiving a placebo to a group that is receiving nothing
  - Minimize investigator bias in assessing
    - adverse events
    - treatment outcomes

32

Logistical Issues

- Blinding is not always possible
  - Placebo not always possible to be identical in appearance
    - weight of fiber, hardness of calcium,
  - Side effects of treatment may be noticeable
    - skin discoloration with beta-carotene
  - Burden of treatment may not be ethical
    - surgery, hospitalizations

33

Blinded Evaluation

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

34

Issues

- Issues that must be addressed with blinded experiments
  - Appearance of treatments
  - Dosage, administration schedules
  - Coding and dispensing treatments
  - When and how to unblind
    - Emergent situations
    - Only unblind when treatment of toxicities differs between therapies
  - Assessing how well the blind was maintained

35

When Blinding is Unnecessary

- Blinding less of an issue with harder endpoints
  - The more objective the measurement of outcome, the less important blinding is to the scientific credibility of a clinical trial.
  - (Of course, the ideal is a blinded experiment with solidly objective endpoints.)

36

Why Randomize?

Statistical Role of Variables in Analysis

Where am I going?

- Ultimately a RCT is designed to compare outcomes across groups in a statistical analysis
- It is useful to review the components of a statistical analysis model in order to
  - develop a standard nomenclature and
  - discuss the goals and impact of randomization.

37

Second Statistical Refinement

- The group receiving the treatment will tend to have outcome measurements that are

higher than, / lower than, / about the same as

than

an absolute standard, or / measurements in an otherwise comparable group (that did not receive the treatment)

38

Ex: Smoking Effect on FEV

- Scientific question
  - Does smoking lead to lower lung function in kids?
- Study design
  - 654 healthy children
  - Measure smoking by self report
  - Measure lung function by FEV
    - Forced expiratory volume: maximum volume of air that can be exhaled in 1 second

39

The Data

SMOKERS

1.953 2.236 3.428 3.208 1.694 3.957 4.789 2.384 3.074 2.387 3.835 2.599 4.756 3.086 4.309 3.413 2.975 3.169 3.343 3.751 2.216 3.078 3.186 3.297 2.304 3.102 2.677 3.297 3.498 2.759 2.953 3.785 2.276 4.637 3.038 3.120 3.339 3.152 3.104 4.045 4.763 3.069 4.506 3.519 3.688 2.679 2.198 3.345 3.082 2.903 3.004 3.406 3.122 3.330 2.608 3.799 4.086 4.070 2.264 4.404 2.278 4.872 4.270 3.727 2.795

NONSMOKERS

1.708 1.724 1.720 1.558 1.895 2.336 1.919 1.415 1.987 1.942 1.602 1.735 2.193 2.118 2.258 1.932 1.472 1.878 2.352 2.604 1.400 1.256 0.839 2.578 2.988 1.404 2.348 1.755 2.980 2.100 1.282 3.000 2.673 2.093 1.612 2.175 2.725 2.071 1.547 2.004 3.135 2.420 1.776 1.931 1.343 2.076 1.624 1.344 1.650 2.732 2.017 2.797 3.556 1.703 1.634 2.570 3.016 2.419 1.569 1.698 2.123 2.481 1.481 1.577 1.940 1.747 2.069 1.631 1.536 2.560 1.962 2.531 2.715 2.457 2.090 1.789 1.858 1.452 3.842 1.719 2.111 1.695 2.211 1.794 1.917 2.144 1.253 2.659 1.580 2.126 3.029 2.964 1.611 2.215 2.388 2.196 1.751 2.165 1.682 1.523 1.292 1.649 2.588 0.796 2.574 1.979 2.354 1.718 1.742 1.603 2.639 1.829 2.084 2.220 1.473 2.341 1.698 1.196 1.872 2.219 2.420 1.827 1.461 1.338 2.090 1.697 1.562 2.040 1.609 2.458 2.650 1.429 1.675 1.947 2.069 1.572 1.348 2.288 1.773 0.791 1.905 2.463 1.431 2.631 3.114 2.135 1.527 2.293 3.042 2.927 2.665 2.301 2.460 2.592 1.750 1.759 1.536 2.259 2.048 2.571 2.046 1.780 1.552 1.953 2.893 1.713 2.851 1.624 2.631 1.819 1.658 2.158 1.789 3.004 2.503 1.933 2.091 2.316 1.704 1.606 1.165 2.102 2.320 2.230 1.716 1.790 1.146 2.187 2.717 1.796 1.335 2.119 1.666 1.826 2.709 2.871 1.092 2.262 2.104 2.166 1.690 2.973 2.145 1.971 2.095 1.697 2.455 1.920 2.164 2.130 2.993 2.529 1.726 2.442 1.102 2.056 1.808 2.305 1.969 1.556 1.072 2.042 1.512 1.423 3.681 1.991 1.897 1.370 1.338 2.016 2.639 1.389 1.612 2.135 2.681 3.223 1.796 2.010 1.523 1.744 2.485 2.335 1.415 2.076 2.435 1.728 2.850 1.844 1.754 1.343 2.303 2.246 2.476 3.239 2.457 2.382 1.640 1.589 2.056 2.226 1.886 2.833 1.715 2.631 2.550 1.912 1.877 1.935 1.539 2.803 2.923 2.358 2.094 1.855 1.535 2.135 1.930 2.182 1.359 2.002 1.699 2.500 2.366 2.069 1.418 2.333 1.514 1.758 2.535 2.564 2.487 1.591 1.624 2.798 1.691 1.999 1.869 1.004 1.427 1.826 2.688 1.657 1.672 2.015 2.371 2.115 2.328 1.495 2.884 2.328 3.381 2.170 3.470 3.058 1.811 2.524 2.642 3.741 4.336 4.842 4.550 2.841 3.166 3.816 2.561 3.654 2.481 2.665 3.203 3.549 3.222 3.111 3.490 3.147 2.520 2.292 2.889 2.246 1.937 2.646 2.957 4.007 2.386 3.251 2.762 3.011 4.305 3.906 3.583 3.236 3.436 3.058 3.007 3.489 2.864 2.819 2.250 4.683 2.352 3.108 3.994 4.393 2.592 3.193 2.346 3.515 2.754 2.720 2.463 2.633 3.048 3.111 3.745 2.094 3.183 3.977 3.354 3.411 3.171 3.887 2.646 2.504 3.587 3.845 2.971 2.891 1.823 2.417 2.175 2.735 4.273 2.976 4.065 2.318 3.596 3.395 2.751 2.673 2.556 2.542 2.608 2.354 1.458 3.795 2.491 3.060 2.545 2.993 3.305 3.774 2.855 2.988 2.498 3.169 2.887 2.704 3.515 3.425 2.287 2.434 2.365 2.696 2.868 2.813 3.255 4.593 4.111 1.916 1.858 3.350 2.901 2.241 4.225 3.223 5.224 4.073 4.080 2.606 4.411 3.791 3.089 2.465 3.200 2.913 4.877 2.358 3.279 2.581 2.347 2.691 2.827 1.873 2.538 2.758 3.050 3.079 2.201 1.898 3.403 3.501 2.578 1.665 2.081 2.974 4.073 4.448 3.984 2.250 2.752 3.680 2.862 3.023 3.681 3.255 3.692 2.356 4.591 3.082 3.258 2.216 3.247 4.324 2.362 2.563 3.206 3.585 4.720 3.331 5.083 2.417 2.364 2.341 3.231 3.078 3.369 3.529 2.866 2.891 3.022 3.127 2.866 2.605 3.056 2.569 2.501 3.320 2.123 3.780 3.847 3.924 2.132 2.752 2.449 3.456 3.073 2.688 3.329 4.271 3.530 2.928 2.689 2.332 2.934 3.110 2.894 2.435 2.838 3.035 4.831 2.812 2.714 3.086 3.519 4.232 2.770 3.341 3.090 2.531 2.822 2.935 2.568 2.387 2.499 4.130 3.001 3.132 3.577 3.222 3.280 2.659 2.822 2.140 4.203 2.997 2.562 3.082 3.806 2.458 2.391 3.141 2.579 2.100 2.785 4.284 2.906 5.102 4.429 4.279 4.500 2.635 3.082 3.387 5.793 3.985 4.220 4.724 3.731 3.500 3.674 5.633 3.645 2.887 3.960 4.299 2.981 4.504 5.638 2.853 3.211

Ex: Comparison of Groups with T test

- Difference between mean FEV between smokers and nonsmokers
  - | T | = 7.1496 on df = 83.27
  - Two –sided P value: < 0.0001
    - Statistically significant diff between dose groups

41

Ex: Analysis by Dose Group

- Mean FEV by self-reported smoking status group
  - Nonsmokers (N = 589)
    - Average FEV 2.57 l / sec
    - 95% CI: 2.50 l / sec to 2.63 l / sec
  - Smokers (N = 65)
    - Average FEV 3.28 l/ sec
    - 95% CI: 3.09 l / sec to 3.46 l / sec

42

Ex: Comparison of Groups with T test

- Difference in mean FEV between smokers and nonsmokers
  - Point estimate: 0.711 l / sec higher FEV in smokers
  - 95% CI: 0.513 l / sec to 0.908 / sec higher in smokers
    - Above observation is not atypical if true difference in average is between 0.513 and 0.908
  - Two –sided P value: < 0.0001
    - Statistically significant diff between smoking groups

43

GM: Unadjusted Interpretation

- Smoking effect
  - Geometric mean of FEV is 10.8% higher in smokers than in nonsmokers
    - 95% CI: 4.1% to 17.9% higher
    - These results are atypical of what we might expect with no true difference between groups: P = 0.001

44

GM: Age Adjusted Interpretation

- Smoking effect
  - Geometric mean of FEV is 5.0% lower in smokers than in nonsmokers of the same age
    - 95% CI: 12.2% lower to 1.6% higher
    - These results are not atypical of what we might expect with no true difference between groups of the same age: P = 0.136

45

Age, Ht Adjusted Interpretation

- Smoking effect
  - Geometric mean of FEV is 5.2% lower in smokers than in nonsmokers of the same age and height
    - 95% CI: 9.6% to 0.6% lower
    - These results are atypical of what we might expect with no true difference between groups of the same age and height: P = 0.027

46

Ex: Take Home Message

- Our scientific question was not

  Is there a difference between smokers' and nonsmokers' average FEV?

- But rather ("personalized medicine"?)

  Do smokers average lower FEV than otherwise comparable nonsmokers?

47

Real-life Examples

- Effects of arrhythmias post MI on survival
  - Observational studies: high risk for death
  - CAST: 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

48

Statistical Role of Variables

- Outcome (response) variable(s)
  - Primary and surrogates
- Predictor(s) of interest (define main groups)
- Subgroups of interest for effect modification
- Potential confounders
- Variables that add precision to analysis
  - 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

49

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 measure
      - Mean, geometric mean, median, proportion, odds, hazard, etc.
    - Comparison across groups
      - Difference, ratio

50

Ignoring Effect Modification

- By design or 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

51

Adjusted Analyses

- If effect modification exists, an analysis adjusting 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)
- "Intent to Cheat" analysis
  - Gain a larger indication by including strata with no effect

52

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 N than to prove effect modification)

53

Confounding

- Definition of confounding
  - The association between a predictor of interest 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
- (A manifestation of Simpson's paradox)

54

Next

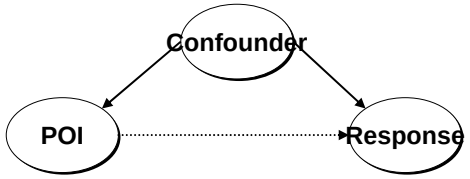
Adjustment for Covariates

- 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

55

Classical Confounder

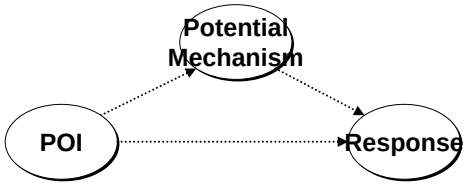
- A clear case of confounding is when some third variable is a "cause" of both the POI and response
  - We generally adjust for such a confounder



56

Causal Pathway

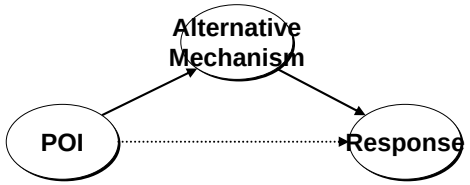
- A variable in the causal pathway of interest is not a confounder
  - We would not adjust for such a variable (lest we lose ability to detect the effect)



57

Causal Pathway

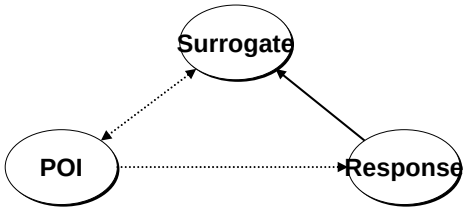
- We would want to adjust for a variable in a causal pathway not of interest
  - E.g., work stress causing ulcers by hormonal effects versus alcoholism



58

Surrogate for Response

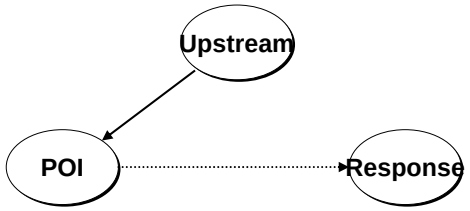
- NOT a confounder: Adjustment for such a variable is a very BAD thing to do



59

"Upstream" Variable

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



60

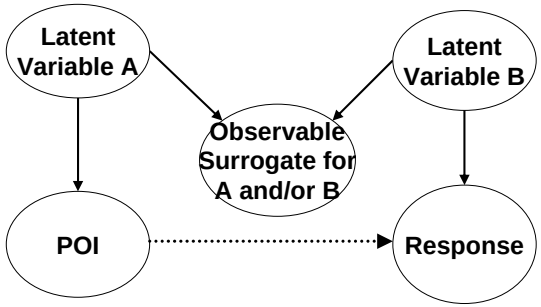
Unadjusted, Adjusted Analyses

- Confounding typically produces a difference between unadjusted and adjusted analyses, but those symptoms are not proof of confounding
  - Such a difference can occur times when there is no confounding
    - "Precision" variables in logistic, PH regression
    - Complicated causal pathways

61

Complicated Causal Pathway

- Adjustment for Variable C would produce a spurious association (effect modification)



62

Precision

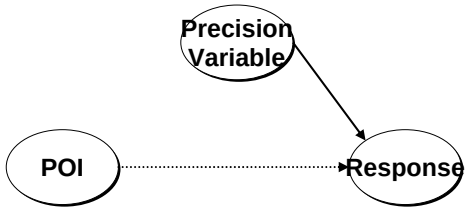
- Sometimes we choose the exact scientific question to be answered on the basis of which question can be answered most precisely
  - In general, questions can be answered more precisely if the within group distribution is less variable
    - Comparing groups that are similar with respect to other important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

63

Next

Precision Variable

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



64

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

65

Why Randomize?

Common Statistical Analysis Models

Where am I going?

- The scientific question posed by a clinical trial is typically translated into a statistical comparison of probability distributions
  - Unadjusted or adjusted comparison of summary measures
- We will need to describe the statistical implications of any randomization strategy in the context of statistical analysis model
  - Notation for regression on means, odds, or hazards

66

Summary Measures

- The measures commonly used to summarize and compare distributions vary according to the types of data
  - Means: binary; quantitative
  - Medians: ordered; quantitative; censored
  - Proportions: binary; nominal
  - Odds: binary; nominal
  - Hazards: censored
    - hazard = instantaneous rate of failure

67

"Everything is Regression"

- The most commonly used two sample tests are special cases of regression
  - Regression with a binary predictor
    - Linear → t test
    - Logistic → chi square (score test)
    - Proportional hazards → logrank (score test)

68

General Regression

- General notation for variables and parameter
  - Y_i Response measured on the *i*th subject
  - X_i Value of the POI for the *i*th subject
  - $W_{1i}, W_{2i}, \ldots$ Value of adjustment variables for the *i*th subject
  - θ_i Parameter of distribution of Y_i
  - The parameter might be the mean, geometric mean, odds, rate, instantaneous risk of an event (hazard), etc.

69

Multiple Regression

- General notation for simple regression model
  - $g(\theta_i) = \beta_0 + \beta_1 \times X_i + \beta_2 \times W_{1i} + \beta_3 \times W_{2i} + \cdots$
  - $g(\ )$ "link" function used for modeling
  - β_0 "Intercept"
  - β_1 "Slope for Pred of Interest X)"
  - β_j "Slope for covariate W_{j-1}"
  - The link function is usually either none (means) or log (geom mean, odds, hazard)

70

Borrowing Information

- Use other groups to make estimates in groups with sparse data
  - Intuitively: 67 and 69 year olds would provide some relevant information about 68 year olds
  - Assuming straight line relationship tells us how to adjust data from other (even more distant) age groups
    - If we do not know about the exact functional relationship, we might want to borrow information only close to each group

71

Defining "Contrasts"

- Define a comparison across groups to use when answering scientific question
  - If straight line relationship in parameter, slope for POI is difference in parameter between groups differing by 1 unit in X when all other covariates in model are equal
  - If nonlinear relationship in parameter, slope is average difference in parameter between groups differing by 1 unit in X "holding covariates constant"
    - Statistical jargon: a "contrast" across the groups

72

Regression Models

- According to the parameter compared across groups
  - Means → Linear regression
  - Geom Means → Linear regression on logs
  - Odds → Logistic regression
  - Rates → Poisson regression
  - Hazards → Proportional Hazards regr
  - Quantiles → Parametric survival regr

73

Comparison of Models

- The major difference between regression models is interpretation of the parameters
  - Summary: Mean, geometric mean, odds, hazards
  - Comparison of groups: Difference, ratio
- Issues related to inclusion of covariates remain the same
  - Address the scientific question
    - Predictor of interest; Effect modifiers
  - Address confounding
  - Increase precision

74

Interpretation of Parameters

- Intercept
  - Corresponds to a population with all modeled covariates equal to zero
    - Most often outside range of data; quite often impossible; very rarely of interest by itself
- Slope
  - A comparison between groups differing by 1 unit in corresponding covariate, but agreeing on all other modeled covariates
    - Sometimes impossible to use this definition when modeling interactions or complex curves

75

Points Meriting Repeated Emphasis

- Common regression models allow us to consider both adjusted and unadjusted analyses
- Generally reasonable distribution-free inference
  - Linear regression
    - Extremely robust
  - Logistic regression
    - Some issues with model mis-specification
  - Proportional hazards model
    - Dependence on censoring distribution
    - Robust if log hazard linear in log time over support of censoring distribution

76

Why Randomize?

Four Important Questions of Regression

Where am I going?

- The fundamental statistical distinctions between unadjusted and adjusted regression models are central to the goals of randomization
- We thus want to be able to consider the relationships between
  - unadjusted and adjusted parameters, and
  - the standard errors of the two parameter estimates.

77

Adjustment for Covariates

- We "adjust" for other covariates
  - Define groups according to
    - Predictor of interest, and
    - Other covariates
  - Compare the distribution of response across groups which
    - differ with respect to the Predictor of Interest, but
    - are the same with respect to the other covariates
      - "holding other variables constant"

78

Unadjusted vs Adjusted Models

- Adjustment for covariates changes the scientific question
  - Unadjusted models
    - Slope compares parameters across groups differing by 1 unit in the modeled predictor
      - Groups may also differ with respect to other variables
  - Adjusted models
    - Slope compares parameters across groups differing by 1 unit in the modeled predictor but similar with respect to other modeled covariates

79

Interpretation of Slopes

- Difference in interpretation of slopes

$$\text{Unadjusted Model}: \quad g\left[\theta \,|\, X_i\right] = \beta_0 + \beta_1 \times X_i$$

  - β_1 = Compares θ for groups differing by 1 unit in X
    - (The distribution of W might differ across groups being compared)

$$\text{Adjusted Model}: \; g\left[\theta \,|\, X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$$

  - γ_1 = Compares θ for groups differing by 1 unit in X, but agreeing in their values of W

80

Comparing models

Unadjusted $g\left[\theta \mid X_i\right] = \beta_0 + \beta_1 \times X_i$

Adjusted $g\left[\theta \mid X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$

Science: When is $\gamma_1 = \beta_1$?

When is $\hat{\gamma}_1 = \hat{\beta}_1$?

Statistics: When is $se\left(\hat{\gamma}_1\right) = se\left(\hat{\beta}_1\right)$?

When is $\hat{se}\left(\hat{\gamma}_1\right) = \hat{se}\left(\hat{\beta}_1\right)$?

81

General Results

- These questions can not be answered precisely in the general case
  - However, in linear regression we can derive exact results
  - These will serve as a basis for later examination of
    - Logistic regression
    - Poisson regression
    - Proportional hazards regression

82

Linear Regression

- Difference in interpretation of slopes

Unadjusted Model: $E\left[Y_i \mid X_i\right] = \beta_0 + \beta_1 \times X_i$

  - β_1 = Diff in mean Y for groups differing by 1 unit in X
    - (The distribution of W might differ across groups being compared)

Adjusted Model: $E\left[Y_i \mid X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$

  - γ_1 = Diff in mean Y for groups differing by 1 unit in X, but agreeing in their values of W

83

Relationships: True Slopes

- The slope of the unadjusted model will tend to be

$$\beta_1 = \gamma_1 + \rho_{XW} \frac{\sigma_W}{\sigma_X} \gamma_2$$

- Hence, true adjusted and unadjusted slopes for X are estimating the same quantity only if
  - $\rho_{XW} = 0$ (X and W are truly uncorrelated), OR
  - $\gamma_2 = 0$ (no association between W and Y after adjusting for X)

84

Relationships: Estimated Slopes

- The estimated slope of the unadjusted model will be

$$\hat{\beta}_1 = \hat{\gamma}_1\left(1+\hat{\gamma}_2 r_{XW}\left[\frac{s_W}{s_X\left(r_{YX}-r_{YW}r_{XW}\right)}\right]\right)$$

- Hence, estimated adjusted and unadjusted slopes for X are equal only if
  - $r_{XW} = 0$ (X and W are uncorrelated in the sample, which can be arranged by experimental design), OR
  - $\hat{\gamma}_2 = 0$ (which cannot be predetermined, because Y is random)
  - $s_W = 0$ (W is controlled at a single value in which case $r_{XW} = 0$)

85

Relationships: True SE

Unadjusted Model $\left[se\left(\hat{\beta}_1\right)\right]^2 = \frac{Var\left(Y \mid X\right)}{nVar\left(X\right)}$

Adjusted Model $\left[se\left(\hat{\gamma}_1\right)\right]^2 = \frac{Var\left(Y \mid X,W\right)}{nVar\left(X\right)\left(1-r_{XW}^2\right)}$

$$Var\left(Y \mid X\right) = \gamma_2^2 Var\left(W \mid X\right) + Var\left(Y \mid X,W\right)$$

$$\sigma_{Y|X}^2 = \gamma_2^2\sigma_{W|X}^2 + \sigma_{Y|X,W}^2$$

86

Relationships: True SE

Unadjusted Model $\left[se\left(\hat{\beta}_1\right)\right]^2 = \frac{Var\left(Y \mid X\right)}{nVar\left(X\right)}$

Adjusted Model $\left[se\left(\hat{\gamma}_1\right)\right]^2 = \frac{Var\left(Y \mid X,W\right)}{nVar\left(X\right)\left(1-r_{XW}^2\right)}$

$$Var\left(Y \mid X\right) = \gamma_2^2 Var\left(W \mid X\right) + Var\left(Y \mid X,W\right)$$

Thus, $se\left(\hat{\beta}_1\right) = se\left(\hat{\gamma}_1\right)$ if

$$r_{XW} = 0$$

AND

$$\gamma_2 = 0 \quad \text{OR} \quad Var\left(W \mid X\right) = 0$$

87

Relationships: Estimated SE

Unadjusted Model $\left[s\hat{e}\left(\hat{\beta}_1\right)\right]^2 = \frac{SSE\left(Y \mid X\right)/\left(n-2\right)}{\left(n-1\right)s_X^2}$

Adjusted Model $\left[s\hat{e}\left(\hat{\gamma}_1\right)\right]^2 = \frac{SSE\left(Y \mid X,W\right)/\left(n-3\right)}{\left(n-1\right)s_X^2\left(1-r_{XW}^2\right)}$

$$SSE\left(Y \mid X\right) = \sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i\right)^2$$

$$SSE\left(Y \mid X,W\right) = \sum\left(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 \times X_i - \hat{\gamma}_2 \times W_i\right)^2$$

88

Relationships: Estimated SE

Unadjusted Model $\left[\hat{se}\left(\hat{\beta}_1\right)\right]^2 = \frac{SSE\left(Y \mid X\right)/(n-2)}{(n-1)s_X^2}$

Adjusted Model $\left[\hat{se}\left(\hat{\gamma}_1\right)\right]^2 = \frac{SSE\left(Y \mid X, W\right)/(n-3)}{(n-1)s_X^2\left(1-r_{XW}^2\right)}$

Thus, $\hat{se}\left(\hat{\beta}_1\right) = \hat{se}\left(\hat{\gamma}_1\right)$ if

$r_{XW} = 0$

AND

$SSE\left(Y \mid X\right)/(n-2) = SSE\left(Y \mid X, W\right)/(n-3)$

89

Residual Squared Error

$$SSE\left(Y \mid X\right) = \sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i\right)^2$$

$$SSE\left(Y \mid X, W\right) = \sum\left(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 \times X_i - \hat{\gamma}_2 \times W_i\right)^2$$

When calculated on the same data :

$$SSE\left(Y \mid X\right) \geq SSE\left(Y \mid X, W\right)$$

90

Relationships: Estimated SE

$$SSE\left(Y \mid X\right) = \sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i\right)^2$$

$$SSE\left(Y \mid X, W\right) = \sum\left(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 \times X_i - \hat{\gamma}_2 \times W_i\right)^2$$

Now $\hat{\beta}_1 = \hat{\gamma}_1$ if

$\hat{\gamma}_2 = 0,$ in which case $SSE\left(Y \mid X\right) = SSE\left(Y \mid X, W\right)$

OR

$r_{XW} = 0,$ and $SSE\left(Y \mid X\right) > SSE\left(Y \mid X, W\right)$ if $\hat{\gamma}_2 \neq 0$

91

Special Cases

- Behavior of unadjusted and adjusted models according to whether
  - *X* and *W* are uncorrelated (no association in means)
  - *W* is associated with *Y* after adjustment for *X*

	$r_{XW} = 0$	$r_{XW} \neq 0$
$\gamma_2 \neq 0$	Precision	Confounding
$\gamma_2 = 0$	Irrelevant	Var Inflation

92

Simulations

- Unadjusted and adjusted estimates of treatment effect as a function of
  - Effect of a third covariate on mean outcome
  - Association between third covariate and treatment
    - Difference in mean covariate
    - Difference in median covariate

$$\text{Sampling} \quad : \quad E\left[W_i \,\middle|\, X_i\right] = \alpha_0 + \alpha_1 \times X_i$$

$$\text{Unadjusted} : \quad g\left[\theta \,\middle|\, X_i\right] = \beta_0 + \beta_1 \times X_i$$

$$\text{Adjusted} \quad : \; g\left[\theta \,\middle|\, X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$$

93

Linear Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{xw}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.19)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	1.0	1.0 (0.28)	1.0 (0.20)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.28)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.3 (0.29)	0.0 (0.21)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.22)

94

Logistic Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{xw}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.42)	0.0 (0.42)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.40)	0.0 (0.42)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.42)	0.0 (0.43)
Precision	0.0	0.0	0.00	1.0	1.0	0.8 (0.43)	1.0 (0.49)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.43)	0.0 (0.48)
Confound	0.0	0.3	0.15	1.0	0.0	0.2 (0.41)	0.0 (0.47)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.41)	0.0 (0.47)

95

Proportional Hazards Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{xw}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.22)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.21)
Precision	0.0	0.0	0.00	1.0	1.0	0.7 (0.21)	1.0 (0.22)
Confound	0.3	0.3	0.15	1.0	0.0	0.2 (0.21)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.1 (0.20)	0.0 (0.22)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.23)

96

Next

Precision: Linear Regression

- E.g., X, W independent in population (or completely randomized experiment) AND W associated with Y independent of X

$\rho_{XW} = 0 \qquad \gamma_2 \neq 0$

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 \approx \hat{\gamma}_1$
Std Errs	$se(\hat{\beta}_1) > se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) > s\hat{e}(\hat{\gamma}_1)$

97

Precision: Logistic Regression

- Adjusting for a precision variable
  - Deattenuates slope away from the null
  - Standard errors reflect mean-variance relationship
    - Substantially increased power only in extreme cases
      - » (OR > 5 for equal samples sizes of binary W)

		True Value	Estimates
Slopes	$\beta_1 > 0$:	$\beta_1 < \gamma_1$	$\hat{\beta}_1 < \hat{\gamma}_1$
	$\beta_1 < 0$:	$\beta_1 > \gamma_1$	$\hat{\beta}_1 < \hat{\gamma}_1$
Std Errs		$se(\hat{\beta}_1) < se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) < s\hat{e}(\hat{\gamma}_1)$

98

Precision: Poisson Regression

- Adjusting for a precision variable
  - No effect on the slope (similar to linear regression)
    - log ratios are linear in log means
  - Standard errors reflect mean-variance relationship
    - Virtually no effect on power

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 \approx \hat{\gamma}_1$
Std Errs	$se(\hat{\beta}_1) \approx se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) \approx s\hat{e}(\hat{\gamma}_1)$

99

Precision: PH Regression

- Adjusting for a precision variable
  - Deattenuates slope away from the null
  - Standard errors stay fairly constant
    - (Complicated result of binomial mean-variance)

		True Value	Estimates
Slopes	$\beta_1 > 0$:	$\beta_1 < \gamma_1$	$\hat{\beta}_1 < \hat{\gamma}_1$
	$\beta_1 < 0$:	$\beta_1 > \gamma_1$	$\hat{\beta}_1 > \hat{\gamma}_1$
Std Errs		$se(\hat{\beta}_1) \approx se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) \approx s\hat{e}(\hat{\gamma}_1)$

100

Lin Reg: Stratified Randomization

- Stratified (orthogonal) randomization in a designed experiment

$r_{XW} = 0 \qquad \gamma_2 \neq 0$

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 = \hat{\gamma}_1$
Std Errs	$se(\hat{\beta}_1) = se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) > s\hat{e}(\hat{\gamma}_1)$

101

Confounding: Linear Regression

- Causally associated with response and associated with POI in sample

$r_{XW} \neq 0 \qquad \gamma_2 \neq 0$

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1 + \rho_{XW}\frac{\sigma_X}{\sigma_W}\gamma_2$	$\hat{\beta}_1 = \hat{\gamma}_1\left(1 + \hat{\gamma}_2 r_{XW}\left[\frac{s_W}{s_X(r_{YX} - r_{YW}r_{XW})}\right]\right)$
Std Errs	$se(\hat{\beta}_1) \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} s\hat{e}(\hat{\gamma}_1)$

102

Relationships: True SE

Unadjusted Model $\quad \left[se(\hat{\beta}_1)\right]^2 = \frac{Var(Y \mid X)}{nVar(X)}$

Adjusted Model $\quad \left[se(\hat{\gamma}_1)\right]^2 = \frac{Var(Y \mid X, W)}{nVar(X)(1 - r_{XW}^2)}$

$$Var(Y \mid X) = \gamma_2^2 Var(W \mid X) + Var(Y \mid X, W)$$

$$\sigma_{Y|X}^2 = \gamma_2^2 \sigma_{W|X}^2 + \sigma_{Y|X,W}^2$$

103

Confounding: Other Regression

- With logistic, Poisson, PH regression we cannot write down a formula, but
  - As with linear regression, anything can happen

	True Value	Estimates
Slopes	$\beta_1 \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \gamma_2$	$\hat{\beta}_1 \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \hat{\gamma}_1$
Std Errs	$se(\hat{\beta}_1) \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} s\hat{e}(\hat{\gamma}_1)$

104

Variance Inflation

- Associated with POI in sample, but not associated with response

$r_{XW} \neq 0 \qquad \gamma_2 = 0$

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 = \hat{\gamma}_1\left(1 + \hat{\gamma}_2 r_{XW}\left[\frac{s_W}{s_X\left(r_{YX} - r_{YW} r_{XW}\right)}\right]\right)$
Std Errs	$se\left(\hat{\beta}_1\right) < se(\hat{\gamma}_1)$	$s\hat{e}\left(\hat{\beta}_1\right) < s\hat{e}(\hat{\gamma}_1)$

105

Var Inflation: Other Regressions

- With logistic, Poisson, PH regression we cannot write down a formula, but
  - Similar to linear regression

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \hat{\gamma}_1$
Std Errs	$se\left(\hat{\beta}_1\right) < se(\hat{\gamma}_1)$	$s\hat{e}\left(\hat{\beta}_1\right) < s\hat{e}(\hat{\gamma}_1)$

106

Irrelevant Variables

- Uncorrelated with POI in sample, and not associated with response
  - Slight loss of precision in all regressions

$r_{XW} = 0 \qquad \gamma_2 = 0$

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 = \hat{\gamma}_1$
Std Errs	$se\left(\hat{\beta}_1\right) = se(\hat{\gamma}_1)$	$s\hat{e}\left(\hat{\beta}_1\right) < s\hat{e}(\hat{\gamma}_1)$

107