Epi 265: Epi Methods 3/ Research Methods in Chronic Disease Epi

Course outline

1. Causal inference from observational data and threats to validity.
2. Analyses with clustered Data: Neighborhoods
3. Data sources: finding what's available, and evaluating tradeoffs
4. Clustered data: Repeated measures
5. Putting together data sources, clustered data, and interactions: evaluating theoretically motivated hypotheses, e.g., lifecourse effects
6. Mediation and Effect Decomposition
7. Selection bias
8. Evaluating Uncertainty: p-values and Cis
9. Power calculations and Publication bias
10. Sampling ?

Student work

- The course will involve two hours of class time per week, with accompanying reading, peer-review comments, take-home quizzes, and a final project.
- Grades for the class will be based on the following:
  - Seven reading responses. Responses are due each Sunday the evening (8pm) before class, but you may skip 1 week. Please post your responses on the class website. Please let me know in advance the week you are skipping. (25% of the grade).
  - Three take-home quizzes (10% of the grade each)
  - Final research project write-up (30%)
  - Feedback to other students, in class or on the forum (15% of the grade)
- For students who wish to have an applied data analysis experience, there is an optional 1-unit independent study. I have posted assignments each week corresponding with completing this IS.
- Most days, there will be an in-class quiz. This quiz is not graded, but you are asked to turn it in at the end of class.
- **Reading response** topics are listed on the CLE site and should be posted on the class website forum for everyone to review. **You are welcome and encouraged to discuss these with classmates (or anyone who will listen). Part of the class grade depends on commenting on one another's reading responses.**

What Kinds of Questions Does Epidemiology Answer?

- Y occur?
- **Prediction**
  - Do A, B, and C predict presence or occurrence of Y? e.g., diagnosis or prognosis
- **Causation**
  - **Total effect**
    - What would changing some exposure – any biological, behavioral, environmental, social (etc.) factor – do to the incidence or prevalence of the health outcome in a population?
    - Will intervening upon X change occurrence of Y?
  - **Public health importance of effect/ attribution**
    - What fraction or how many cases of disease Y can be eliminated if a causal exposure X is eliminated or reduced? This reflects both the effect of X and the prevalence of X.
  - **Mediation**
    - Understanding the mechanisms of causation
    - How does X cause Y?
- **Interaction- which spans prediction and causation**
  - When and for whom does X cause Y? Are the effects different for different kinds of people or in different settings?

Themes

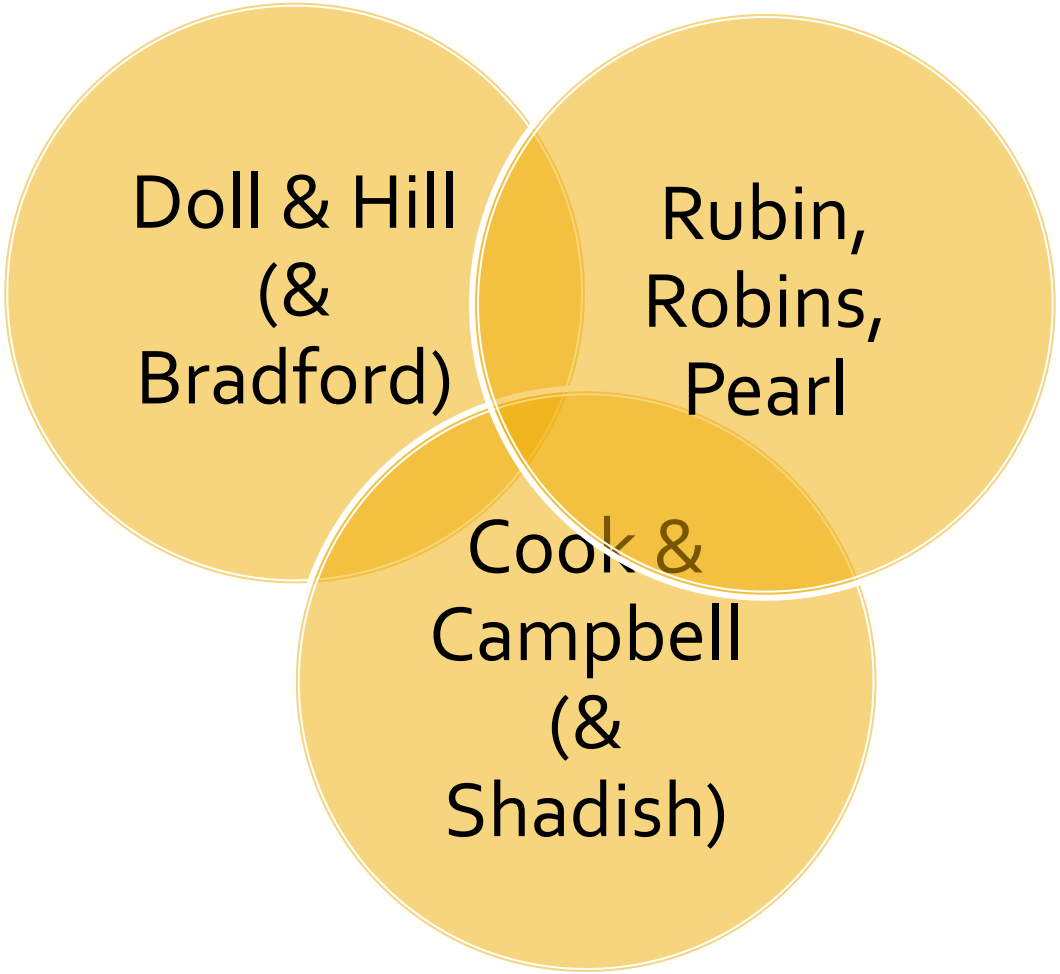
- Insights from other disciplines, especially labor economics
- Linking theoretical questions and frames to the statistical models
- The dynamic feedback between the data, the tools, and the substantive question
- Humility.
- Most ideas are not specific to chronic disease, but the applications will mostly be chronic and some themes emerge that are specific to chronic disease
- Applied papers are *illustrations*: focus on the methods and the correspondence between the research question and the methods.

Announcements

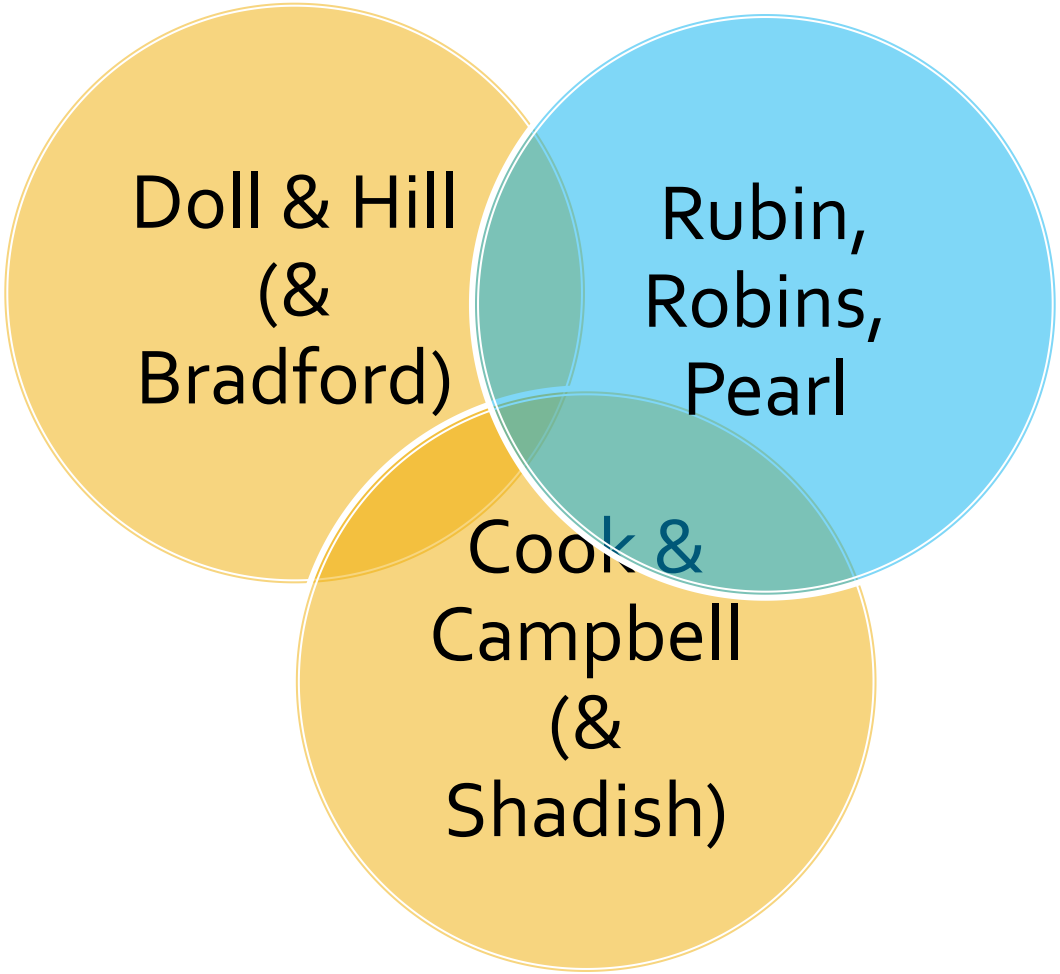
- 13th
- Betz Halloran
- Sam Harper
- Others?

Causal Inference from Observational Data

Alternative traditions for discussion causality



Alternative traditions for discussion causality



When is causal inference the goal?

- When you wish to plan or evaluate an intervention, treatment, or exposure you need to make inferences about the causal effects of that intervention, treatment or exposure.
- The data you happen to be using, and the design of the study, are not relevant to whether your goal is causal inference.
- The data and design determine whether the assumptions required for causal inference are plausible.
- Do not be scared to talk about your *goals*, but be honest about your assumptions and whether they are plausible

Counterfactuals or Potential Outcomes

- If Arlo lives in poverty, he will develop diabetes before age 60.
  - $Y_{X=1}=1$
- If Arlo doesn't live in poverty, he will *not* get diabetes before age 60.
  - $Y_{X=0}=0$
  - The effect of poverty on diabetes is: $Y_{X=1}-Y_{X=0}$
  - Define this for individuals: $Y_{i,X=1}-Y_{i,X=0}$

Arlo actually does live in poverty and he actually does develop diabetes. We never get to see what would have happened to him if he'd avoided poverty.

This is the *fundamental problem of causal inference.*

- Focus on effect for the whole population: $E(Y_{X=1})-E(Y_{X=0})$

Estimating counterfactual values from observed values

- Want to estimate: $E(Y_{X=1})$-$E(Y_{X=0})$
- Estimate: $E(Y_{X=1})$ using $E(Y_{X=1}|X=1)$
- Estimate: $E(Y_{X=0})$ using $E(Y_{X=0}|X=0)$
- So effect estimate is: $E(Y_{X=1}|X=1)$ - $E(Y_{X=0}|X=0)$
- [Note we are ignoring the question of ratio or difference measures for now – difference measures are just more intuitive so we start there]
- Must assume that $E(Y_{X=1}|X=1)=E(Y_{X=1}|X=0)$
- And that $E(Y_{X=0}|X=0)=E(Y_{X=0}|X=1)$
- This is implied by the assumption that X is independent of $\mathbf{Y_X}$

Estimating counterfactual values from observed values

- Population Average Treatment Effect: $E(Y_{X=1})-E(Y_{X=0})$
- Effect of Treatment on the Treated: $E(Y_{X=1}|X=1)-E(Y_{X=0}|X=1)$
- **How would you estimate this?**
- **Do you need to assume:** X is independent of $\mathbf{Y_X}$ **?**

When are effects identifiable?

Instead we observe the diabetes status of Jeb, who's a lot like Arlo but avoided poverty. Jeb did not develop diabetes. We assume that Arlo and Jeb are "exchangeable", and conclude that Arlo developed diabetes because of his poverty.

We observe the statistical association between poverty and diabetes, and hope that the diabetic outcomes of people who were not impoverished represent the outcomes people who were impoverished **would have had** if they hadn't been poor.

"Confounding is present if our substitute imperfectly represents what our target would have been like under the counterfactual condition."

- Maldonado and Greenland (2002)

Inferring Causation from Association

"Confounding is present if our substitute imperfectly represents what our target would have been like under the counterfactual condition."

- Maldonado and Greenland (2002)

- For any analysis, ask: What is the identifying contrast? Which groups of people am I using to represent one another's counterfactual outcomes?

- What causal structures could have generated the statistical associations we observe? We are interested in knowing whether X caused Y – what are the range of causal structure consistent with the observed statistical associations?

Stable Unit Treatment Value Assumption

- SUTVA:
  - Treatment received by person *i* does not affect response of person *j*, i.e., no interference
  - Violations?

Shadish, Cook and Campbell: influential in social science quantitative research

Don Campbell: "psychologist" died in 1996, at northwestern much of career

Tom Cook: at Northwestern since 1968, sociologist, "commnc'ns research" first jobs in psych departments.

William Shadish: trained at Northwestern in the 1970s, clinical psych/ stats, now at U of C Merced

What kinds of questions do they ask?

- How is children's academic performance influenced by their friend's academic performance?
- How do neighborhoods influence adolescent development?
- How did the Reagan administration crackdown on fraud in the Food Stamp program affect the program's impact?

What are some features of these questions?

- Randomization difficult
- Data are clustered, i.e., observations are not independent
- Relevant constructs often latent, hard to measure
- Level of exposure measure and outcome measure is not the same: exposures are sometimes policies- not features of an individual

Types of Validity (Cook & Campbell Framework)

Have focused more systematically on when you can get the "right" answer with observational data

Implicit tension: using the most "rigorous" approach also means you won't be able to address some of the most important questions.

Important to know if you can get the right answer with more feasible designs.

Cook and Campbell consider various ways you can get the wrong answer in observational research, depending on the study design.

Note ambiguity in terms "observational" "quasi-experimental" or "natural experiments"

Types of Validity (Cook & Campbell Framework)

1. Statistical Conclusion Validity: validity of inferences about statistical associations between treatment and outcome

Types of Validity (Cook & Campbell Framework)

2. Internal validity: validity of inferences about the causal relationship from treatment as measured to outcome as measured within this sample.

Types of Validity (Cook & Campbell Framework)

3. Construct validity: the validity of inferences about the higher order constructs that you thought you were measuring

Types of Validity (Cook & Campbell Framework)

4. External validity: the validity of inferences about whether the cause-effect relationship holds in other populations, other times for the same population, slight variations on treatments or slight variations in measurements.

(note recent work on external validity distinguishes between settings where your data are a subset of the target population and settings where you are inferring to a whole new population)

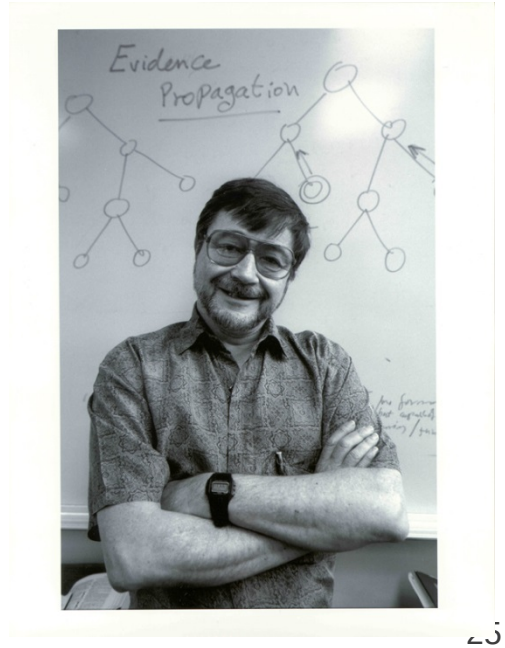
"Causal Inference Methods" in epidemiology dominated by Robins (occ health MD), Pearl (computer scientist) and trainees

Work in causal inference in epidemiology largely corresponds with the Cook & Campbell typology and discussion.

Biostatistics usually focuses on what C & C call statistical conclusion validity.

The majority of emphasis in epi has been on internal validity, with recent attention to external validity.

Many of the threats to (internal) validity discussed by C & C correspond to an arrow on a causal diagram.



Linking to the Cook & Campbell Framework: Statistical Conclusion Validity

Definition

Statistical Conclusion Validity (SCV)

- Refers to the validity with which statements can be made about whether there is a *statistical relationship* between one variable and another.

- This validity does not infer anything about the causality between the two variables, just whether they are associated (i.e., it gets at the 'covariation' component of causal inference)

What Affects SCV

- Anything that increases the probability that you will make a Type 1 error or a Type 2 error will reduce your statistical conclusion validity.
  - Type 1 error: rejecting the null hypothesis when it is true
  - Type 2 error: failing to reject the null hypothesis when it is false.
  - (Type 3 error not discussed: asking the wrong question)
- Anything which decreases your power (increases the probability of a Type II error) will also reduce your statistical conclusion validity.

Threats to Statistical Conclusion Validity

- Low statistical power
- Violated assumptions of statistical tests
- "Fishing" (multiple comparisons)
- Unreliability of measures
- Restriction of range
- Unreliability of treatment implementation
- Extraneous variance in the experimental setting
- Heterogeneity of units
- Inaccurate effect size estimation

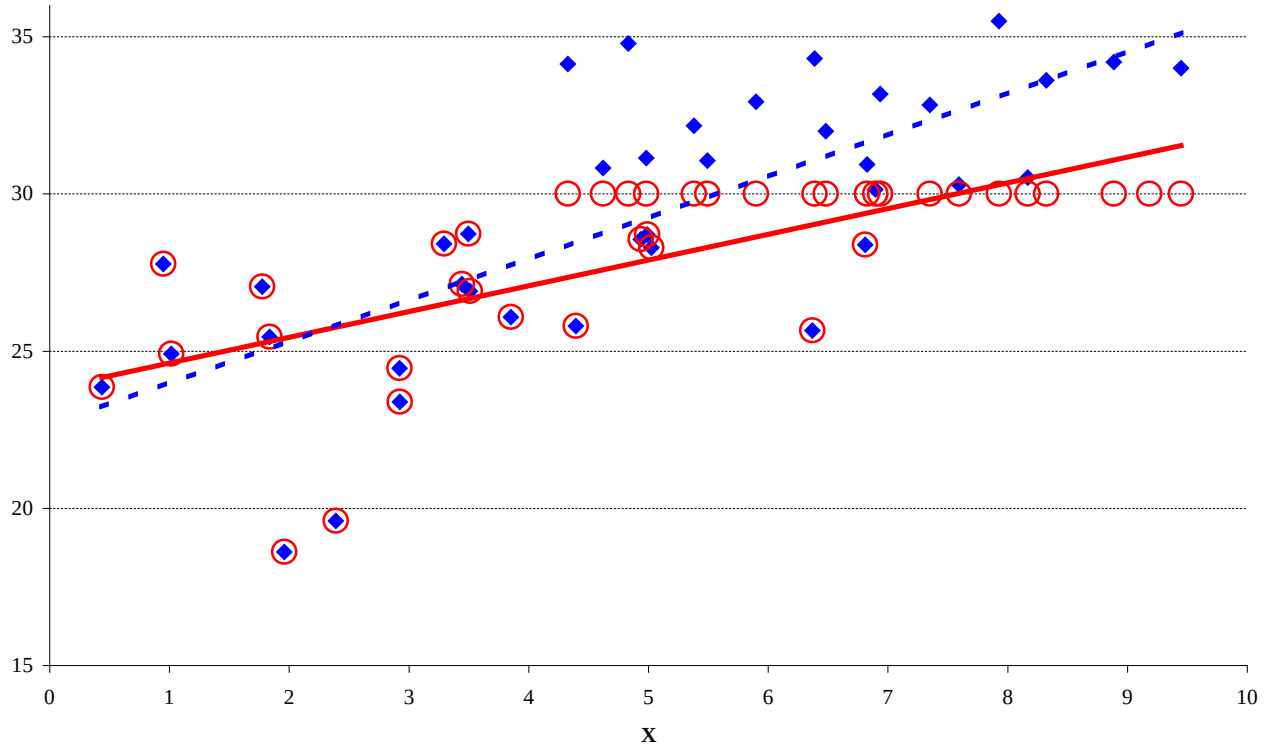
Threats to Statistical Conclusion Validity: Examples and opportunities to address?

- Low statistical power
- Violated assumptions of statistical tests
- "Fishing" (multiple comparisons)
- Unreliability of measures
- Restriction of range
- Unreliability of treatment implementation
- Extraneous variance in the experimental setting
- Heterogeneity of units
- Inaccurate effect size estimation

Low Statistical Power: Examples and opportunities to address?

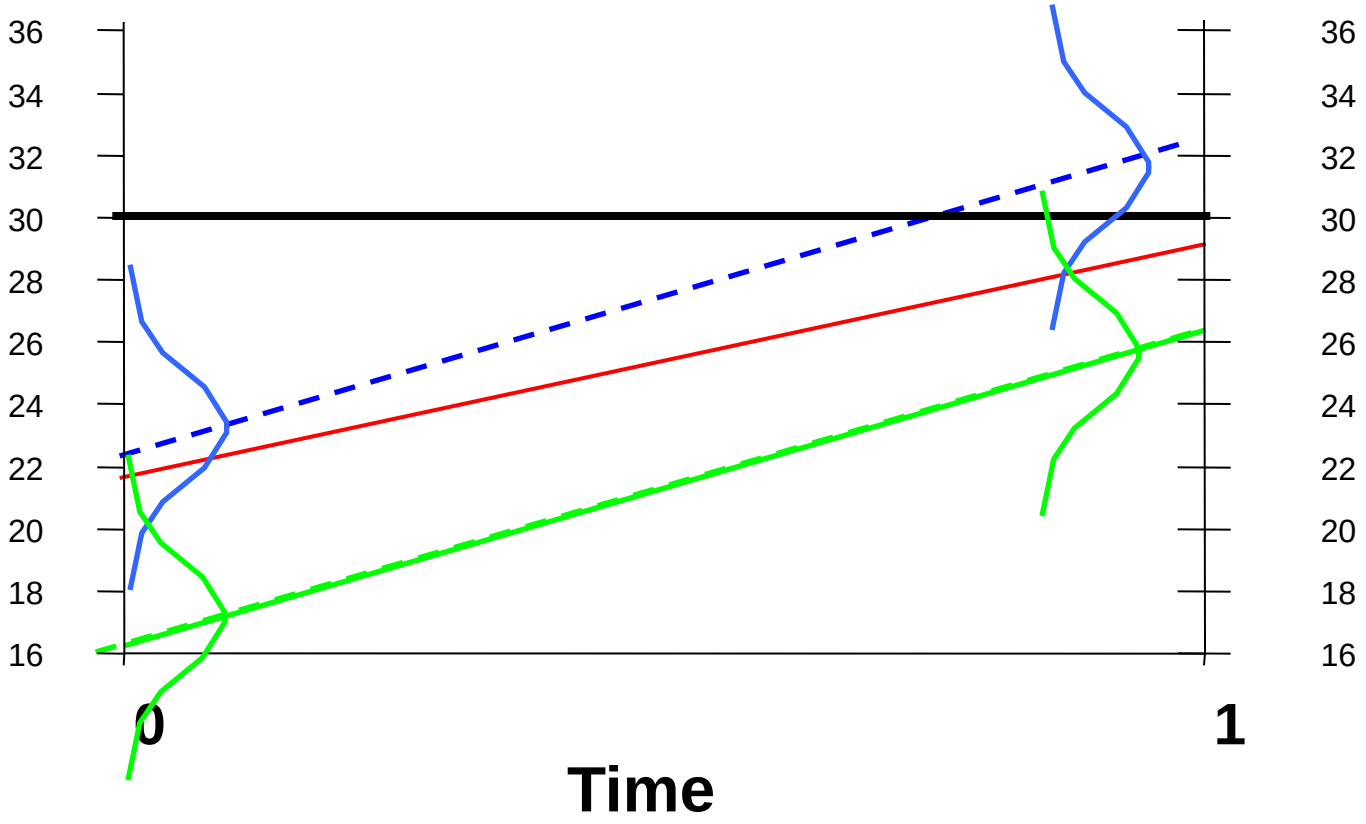
- Type 1 error rate: probability of seeing an association as extreme as this under the null
- Type 2 error rate: probability of failing to detect an association of specified magnitude with given design (sometimes called β)
- Sample size: number of *independent* observations
  - In a clustered sample: increase # of clusters or the # of people in a cluster?
- Magnitude of association (can be conceptualized in terms of standard deviations of the outcome)
- For example, for a comparison of means:
- $N=(z_{\alpha/2}+z_{\beta})^2(\sigma_0^2+\sigma_1^2)/(\mu_0-\mu_1)^2$ to achieve power of $(1-\beta)$ w/ type 1 error rate of α

Restriction of range: Measurement Ceilings



A ceiling on the dependent variable will bias the regression coefficient away from the coefficient for the true outcome variable. Cross sectional: bias is towards the null.

Ceilings in Longitudinal: Unpredictable Bias



With longitudinal data in which some people are improving and others declining, bias is uncertain.

Internal Validity

Definition

Internal Validity

- Refers to the validity with which statements can be made about whether there is a causal relationship between one variable and another in the form in which the variables are manipulated or measured
- In other words, given that there is a statistical association (i.e., SCV is met), is it possible that the association is causal?

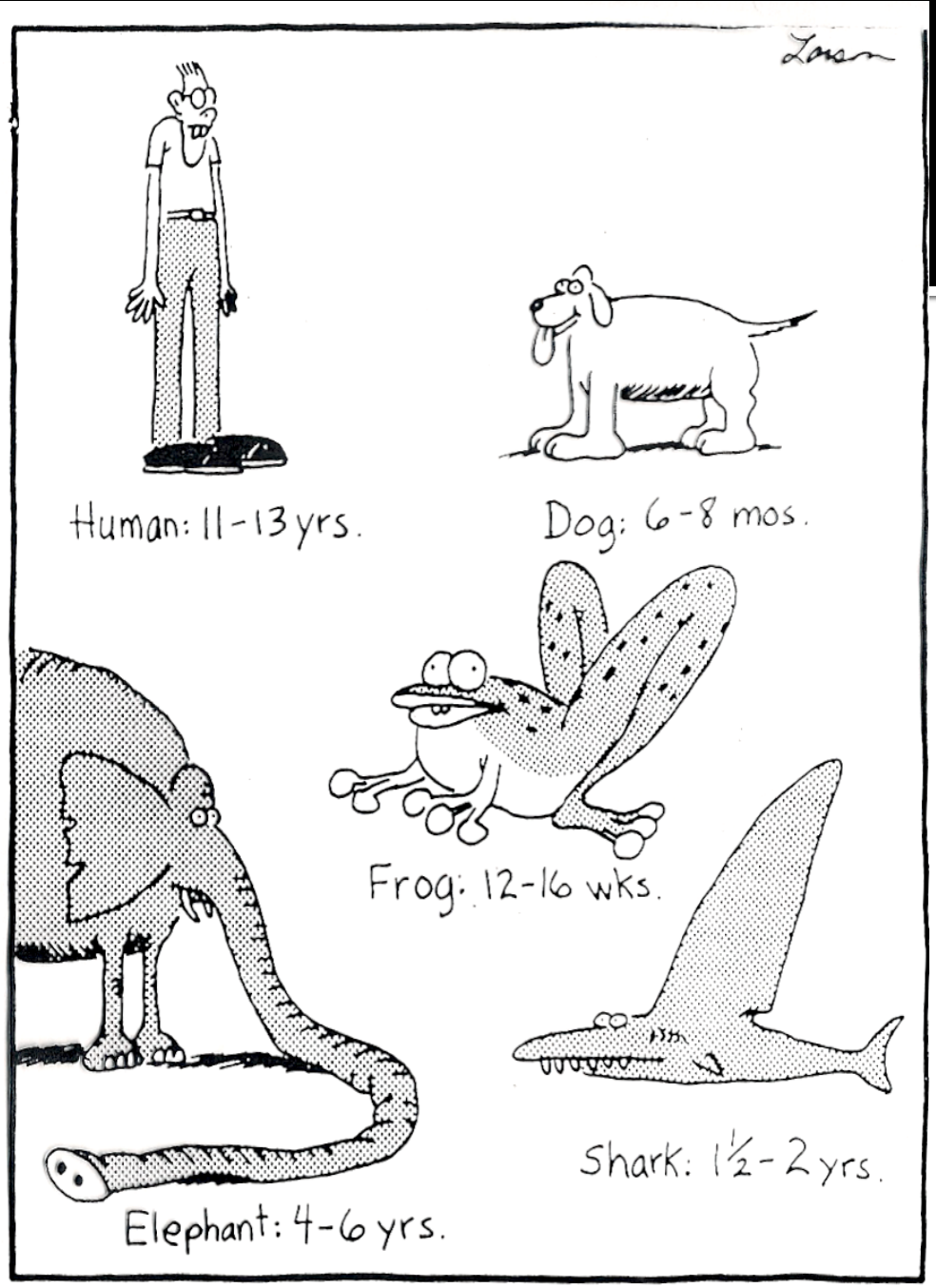
Threats to Internal Validity

Goal is to help you anticipate and rule out, either statistically or in study design.

Sometimes to understand the SCC worries, it helps to think about the *weakest* possible types of studies, for example cross-sectional studies showing a correlation between X and Y or pre-post studies in which you compare outcome Y before and after treatment X.

Threats to Internal Validity

- Ambiguous temporal precedence
- Selection: selection into treatment (this is economist-speak)
- History: Events occurring concurrently could cause the observed effect
- Experimental mortality (attrition)
- Maturation: Naturally occurring changes over time confused with treatment effects



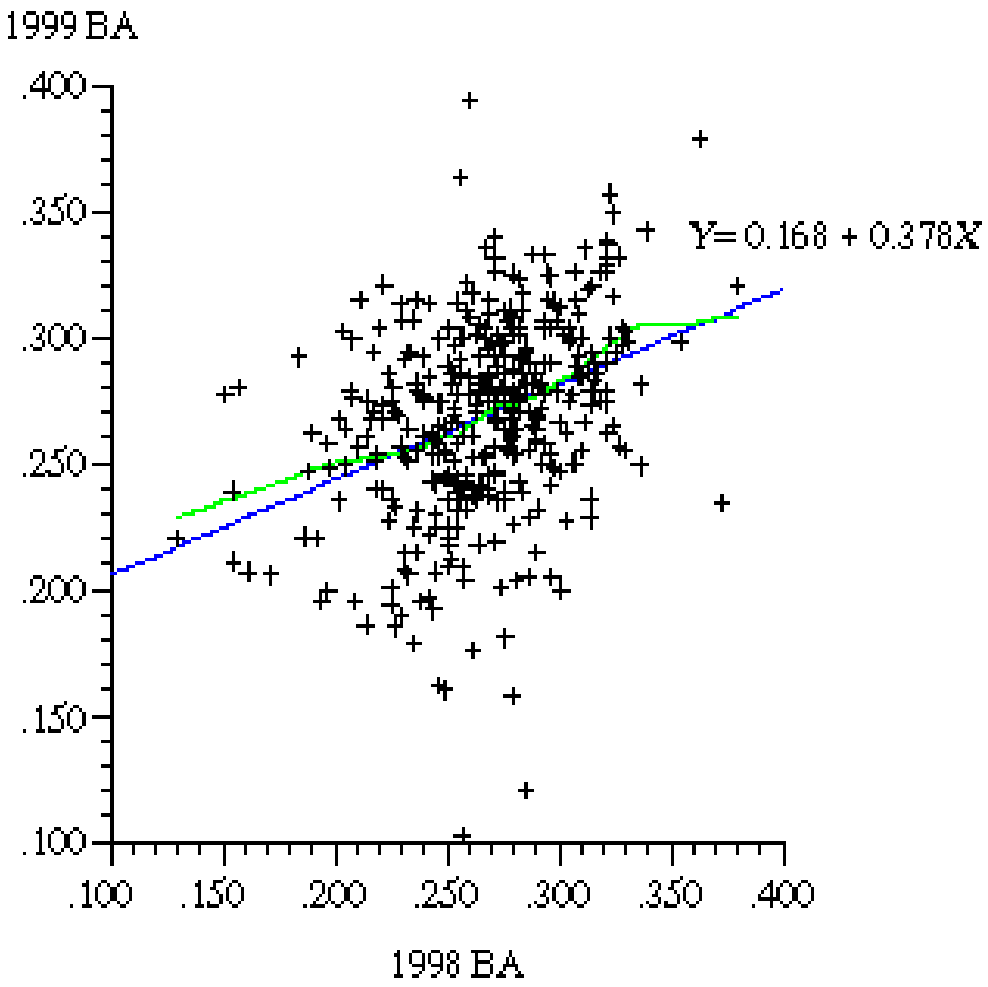
Awkward ages

Threats to Internal Validity Cont'd

- Testing: exposure to a test affects scores on subsequent exposures to the test
- Instrumentation: The measurement tool itself changes
- Statistical regression: Related to maturation: things regress to their mean
- Additive and interactive effects of threats to validity

Statistical Regression

Figure 1. 1998 and 1999 Batting Averages



- Players who do very well in one year tend to "regress" to the population mean in the next.
- Likewise, players who do very badly one year tend to regress towards the population mean the next.
- If BA in 1998 = .400, predict .168+.151=.319 in 1999
- If BA in 1998=100, predict .168+.0378= 172 in 1999

http://www.economics.pomona.edu/GarySmith/Bregress/baseball.html

Threat Level in the Context of Experimental Designs

Temporality	Low -- manipulation
Selection	Low – randomization
History	Low – control group
Experimental mortality (attrition)	Potentially Low to High
Maturation	Low – control group
Testing	Medium
Instrumentation	Medium
Statistical regression	Low – randomization
Contamination	Medium

Evaluate internal validity threats

Does Childbirth Cause Psychiatric Disorders? A Population-based Study Paralleling a Natural Experiment

Munk-Olsen, Trine; Agerbo, Esben

Epidemiology. 26(1):79-84, January 2015.

Background: Childbirth is associated with increased risk of first-time psychiatric episodes, and an unwanted pregnancy has been suggested as a possible etiologic contributor. To what extent childbirth causes psychiatric episodes and whether a planned pregnancy reduces the risk of postpartum psychiatric episodes has not been established.

Methods: We conducted a cohort study using data derived from Danish population registers, including all women having in vitro fertilization (IVF) treatment and their partners with recorded information in the IVF register covering fertility treatments in Denmark at all public and private treatment sites from January 1994 to December 2005. We compared parents and childless persons to examine whether childbirth is directly associated with onset of first-time psychiatric episodes, with incidence rate ratios (risk of first psychiatric inpatient or outpatient treatment) as the main outcome measures.

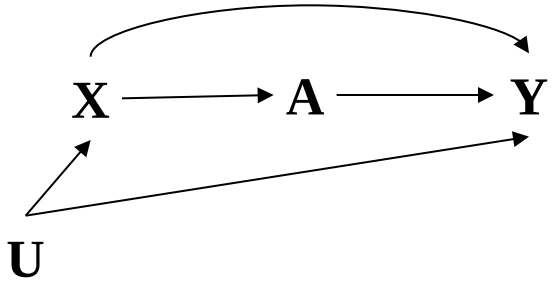
Results: The incidence rate for any type of psychiatric disorder 0 to 90 days postpartum was 11.3 per 1000 person-years (95% confidence interval = 8.2–15.0), and 3.8 (3.4–4.3) among women not giving birth. IVF-treated mothers had an increased risk of a psychiatric episode postpartum (incidence rate ratio [IRR] = 2.9 [2.0–4.2]) compared with the risk of psychiatric episodes in childless women. Risk of psychiatric episodes later than 90 days postpartum was decreased (IRR = 0.9 [0.7–1.0]).

Conclusions: Using a study design paralleling a natural experiment, our results showed that childbirth is associated with first-time psychiatric disorders in new mothers, indicating that a planned pregnancy does not reduce risks of or prevent postpartum psychiatric episodes.

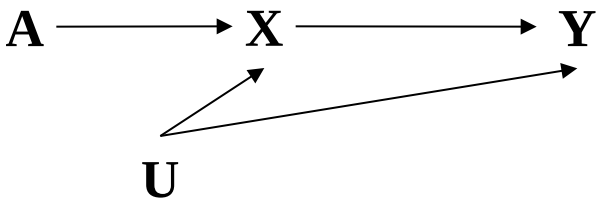
Causal Directed Acyclic Graphs

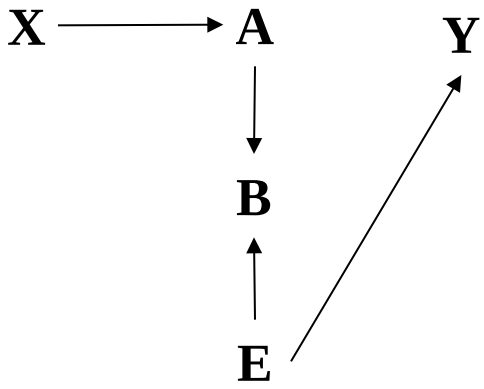
Show your assumptions about the causal relationships among X, Y, and possible covariates in a causal diagram:

•If two variables shown in the graph have a common cause, you must show the cause in the graph.



•Do not allow causal “loops”.





Where do DAGs fit in C & C "types of validity"?

- DAGs help you evaluate internal validity of the association you observe in your sample
- You specify your assumptions about how the world works in the DAG
- You can use this to assess whether, *under these assumptions*, your analysis could have identified the causal effect of interest
- If the world looks like this DAG – would my analysis discover the truth?

Causal Inference Framework vs C & C Types of Validity

Many of the threats to internal validity discussed by C & C correspond to an arrow on a causal diagram.

The discussions by C & C provide a specific story about how those arrows might come to be there.

A key challenge in using DAGs is choosing the right, or a set of plausible, DAGs.

Thinking through the specific stories presented by C & C help draw the DAG corresponding to your research question.

Causal Inference Framework vs C & C Types of Validity

Statistical Conclusion Validity: Rules out chance for statistical associations

Internal Validity: Explanations confounding, reverse causation, or selection/collider bias for statistical associations, evaluated with DAGs / counterfactual framework/ assessment of exchangeability

Construct Validity: Not much attention in causal inference debates (can represent latent variables with DAGs)

External Validity: Less discussion, but one example is the emphasis on a clearly defined intervention corresponding to exposure measure and clarification about irrelevant treatment variants (e.g. would taking the pill with *left* hand change its effects?

How Do the Doll/Hill Criteria Relate?

Doll and Hill criteria would correspond with rules for assessing "internal validity" in C & C's terminology.

- **Strength of Association.**
- **Temporality.**
- **Consistency.**
- **Theoretical Plausibility.**
- **Coherence.**
- **Specificity in the causes.**
- **Dose Response Relationship.**
- **Experimental Evidence.**
- **Analogy.**

How Do the Doll/Hill Criteria Relate?

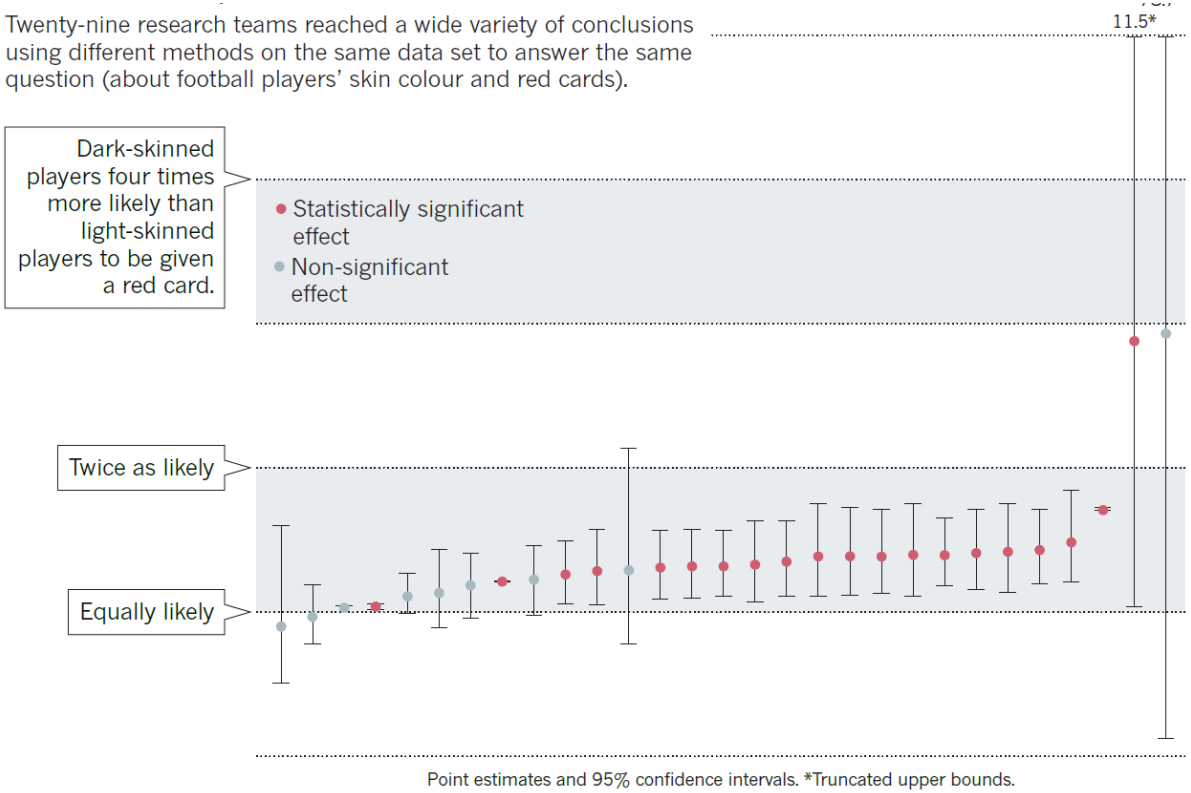
These are all *helpful* to think about, but recognize that only temporality is really necessary. All others are to help make an informed judgment, but they are not "criteria":

- Strength of Association… suggests small biases are unlikely to account for the finding, but some effects are small, or only relevant for a small fraction of the population.
- Consistency… every single person working on a research question may make the same mistakes. This happens all the time. Consistency only increases your confidence if the studies really help rule out one another's weaknesses. Prior consistency reduces the chances that anyone will believe their anomalous results.

Reproducibility/ OSF

- Independent researchers analyzing the same data

(Silberzahn and Uhlmann, Nature 2015)



- New analyses of new data
- New analytic approaches to test the same hypothesis, but relying on different assumptions

How Do the Doll/Hill Criteria Relate?

- Theoretical Plausibility, coherence, and analogy… theory is only as good as what we've already figured out. New discoveries are new because we didn't think that was the way the world (or the body) worked.
- Specificity in the causes… specificity may increase your confidence, but lack of specificity is complete uninformative.
- Dose Response Relationship…lots of risk factors have thresholds.
- Experimental Evidence… the effect is there or not, regardless of whether you have done a trial. It is more convincing if you have experimental evidence though.

Cook Criteria for Comparing Effect Estimates Observational vs RCT

- Experiment and observational study should estimate the same causal quantity: not ATE vs ITT
- Similar populations
- No publication bias
- Both RCT and observational "high quality"
- Evaluate chance as an explanation for divergence of effect estimates

Analyzing RCTs

- ITT: contrast outcomes by randomly assigned treatment groups
- Per protocol: Omit people who didn't adhere to their assigned treatment (this analysis is almost never a good idea)
- As treated: Analyze people based on the treatment they actually received (regardless of assignment (this analysis is rarely a good idea)
- IV: correct the ITT for non-adherence

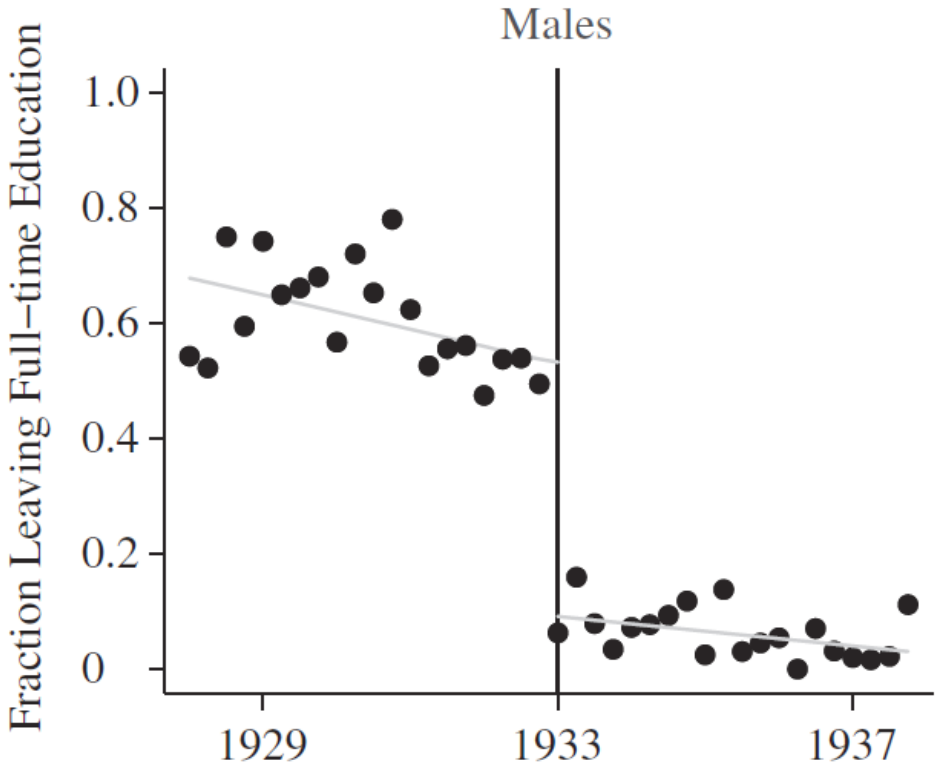
When can you rely on observational studies?

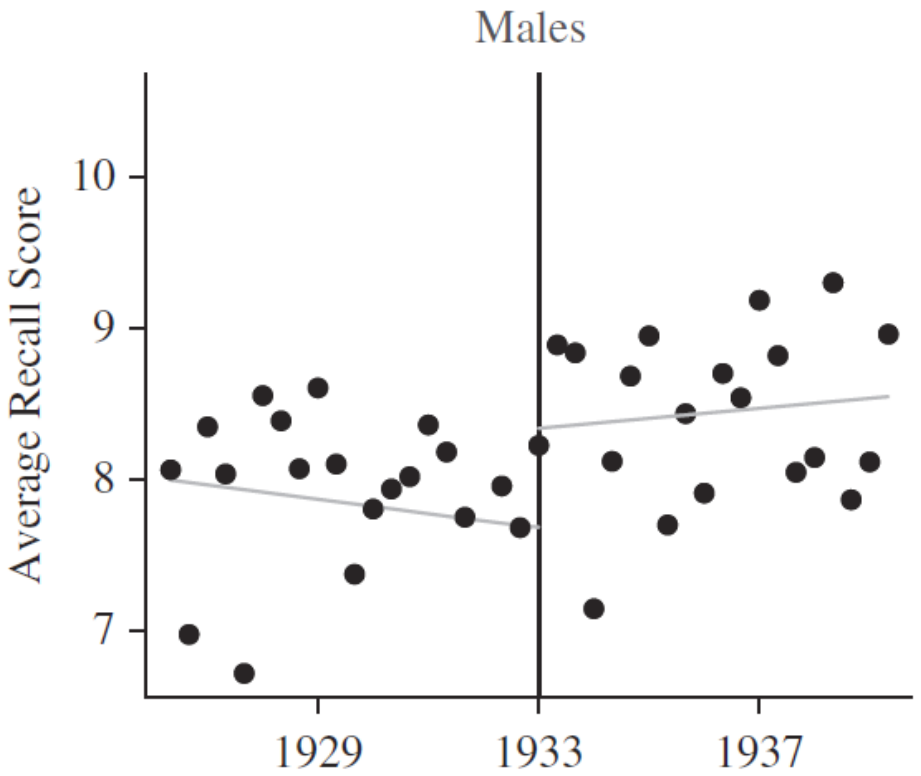
Cook et al argue:

1. Regression discontinuity: compare people right below and right above eligibility cutoff
2. Comparison group defined by similarity to exposed group on pre-exposure value of the outcome variable: note group level emphasis
3. Selection process is well understood

When can you rely on observational studies?

1. Regression discontinuity: compare people right below and right above eligibility cutoff

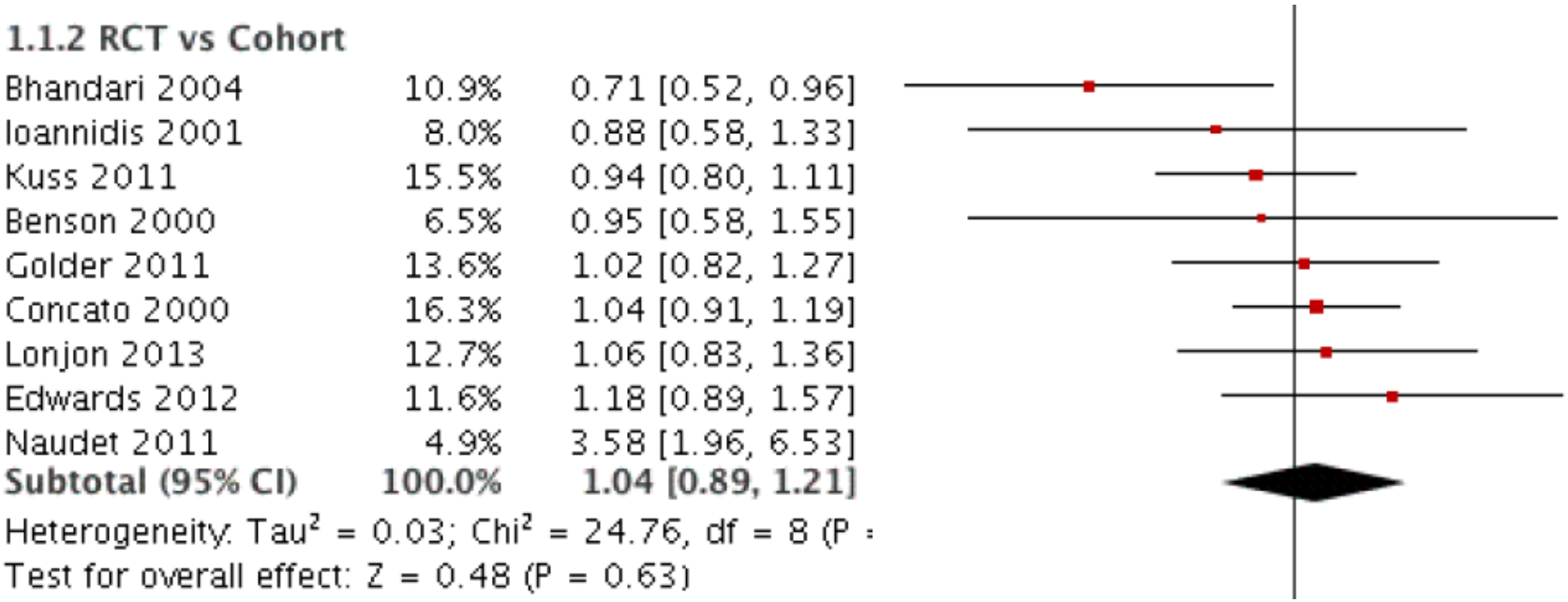




Selection process is well understood? Cochrane Review: OLS vs RCT

"Our results showed that, on average, there is little difference between the results obtained from RCTs and observational studies." Anglemyer, Horvath, and Bero, 2014

Meta-analyzed OR: 1.04 (0.89, 1.21)



End Lecture 1

Inferring Causation From Association

Statistical association between two variables X and Y may be due to:

1. Random fluctuation
2. X caused Y
3. Y caused X
4. X and Y share a common cause
5. The statistical association was induced by conditioning on a common effect of X and Y (as in selection bias).

Random Selection vs Randomization

- Random *selection*: a feature of sample designs: each individual in the target population has a (known) probability of being included in the sample. Does not improve internal validity.
- Random *assignment*: how the exposure is assigned is random and therefore not associated with the potential outcomes.
- Clustered sample: (typically) over-represents people who are similar to each other compared to a simple random sample
- Cluster randomization: assigns everyone in a cluster to the same treatment group and therefore makes people likely to be similar to each other with respect to the outcome (i.e. members of the same cluster) also similar to each other with respect to the exposure.

Selection Selection

- Economists typically use selection to mean something similar to what epidemiologists mean by 'confounding' : selection into treatment
- Epidemiologists use selection to mean selection into or out of the study

Inferring Causation From Association

Statistical association between two variables X and Y may be due to:

1. Random fluctuation
2. X caused Y
3. Y caused X
4. X and Y share a common cause
5. The statistical association was induced by conditioning on a common effect of X and Y (as in selection bias).

How we can use this

- Eliminating four of these explanations is usually the goal of an analysis.
- Knowing these five sources of statistical association helps identify the (set of) causal structure(s) that could have generated the observed statistical associations.