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

Course outline

1. Causal inference from observational data and threats to validity.
   - Threats to validity: integrating perspectives
   - Data sources: finding what's available, and evaluating tradeoffs
   - Sampling: basic
2. Analyses with clustered Data:
   - Cluster randomized trials
   - Neighborhoods as clusters
3. Interactions
   - Review of interaction, effect modification, heterogeneous treatment effects
   - Individuals as clusters (longitudinal data analysis)
4. Evaluating Uncertainty
   - p-values and Confidence Intervals
   - Combining evidence
5. Mediation and Effect Decomposition
6. Power calculations and Publication bias
7. Evaluating theoretically motivated hypotheses

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 Thursday 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 (and all epi PhD students), 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 (with your name) 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?

- **Description**
  - How frequent/common are risk factors/exposures/conditions/diseases?
- **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

- Robins
- Westreich

Causal Inference from Observational Data

Outline Session 1

Section 1: Representative Sampling: basic ideas

Section 2: Causal inference from observational data and threats to validity.

- Threats to validity: integrating perspectives
- Shadish Cook and Campbell framework
  - Statistical conclusion validity
  - Internal validity
  - Construct validity
  - External validity
- Modern causal inference framework (Pearl, Robins)
- Doll/Hill Criteria

Sampling Intro

- Convenience
- **Representative sampling with a sampling frame**
- Representative without a frame
  - Time-location sampling
  - Respondent driven sampling
  - Capture-recapture

Step 1: Who is the target population?

- All humans?
- All people in the US?
- All citizens in the US?
- All females aged 18-39 in the US?
- All non-citizens residing in the US?
- All undocumented citizens residing in the US?

Step 2: Do you need a census, a representative sample, or just some people from the target population?

- You rarely *need* a census.
- To describe a characteristic of the target population, you usually need a representative sample.
  - Representative sample: each element in the population (e.g., each person) has a known, positive probability of selection into the sample.
  - Does *not* require that each person has an equal probability of selection
  - Could be 100% probability of selection
  - Could be .000001% probability of selection
- To evaluate causal effects or even correlations, you may not need a representative sample.

Step 2: Do you need a census, a representative sample, or just some people from the target population?

- You rarely need a census.
- To describe a characteristic of the target population, you usually need a representative sample.
- To evaluate causal effects or even correlations, you may not need a representative sample.
  - What is the average blood pressure of US residents?
  - What is the association between race and blood pressure among US residents?
  - What is the association between blood pressure and stroke among US residents?

Step 3: Can you identify a sampling frame?

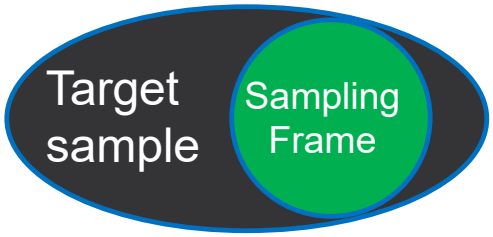
A sampling frame is a list of all eligible elements in the target population.

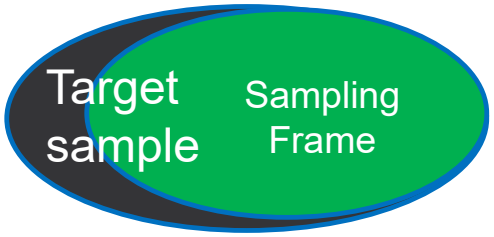
- All humans?
- All people in the US?
- All citizens in the US?
- All voters in San Francisco?
- All females aged 18-39 in San Francisco?
- All non-citizens residing in San Francisco?
- All men who have sex with men residing in SF?
- All people who have been diagnosed as having a stroke in the past two years in SF?

Step 3: Can you identify a sampling frame?

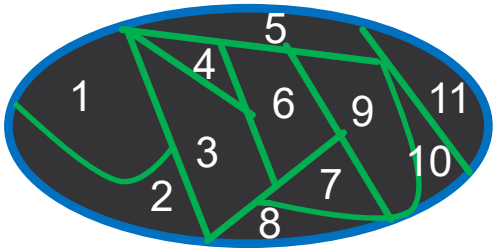
A sampling frame is a list of all eligible elements in the target population.

- Often very tough to identify a perfect sampling frame (i.e., a frame with complete coverage of the target population).
- Is your sampling frame *approximately* complete?





- Can you divide up the population into a list of mutually exclusive groups (clusters) that comprehensively cover the population?
  - All people in the US → 50 US states



Types of Representative Sampling

- Simple Random Sampling
- Systematic Random Sampling
- Stratified Random Sampling
- Multi-stage Sampling
  - Simple Multi-stage Sampling
  - Multi-stage sampling proportionate to size

Types of Representative Sampling

- **Simple Random Sampling:**
  - Each person in sampling frame has equal probability of selection
  - Must start with a list of every element in the target population
- **Systematic Random Sampling:**
  - Decide how many people you want to be in your sample
  - Calculate sampling fraction: sample size/size of whole popln, 1/x
  - Take every x^{th} person from the list of the target population
- **Stratified Random Sampling**
  - Stratify into groups homogeneous on some characteristic
  - Sample with pre-specified probability from each group
- **Multi-stage Sampling**
  - Organize everyone into mutually exclusive groups/clusters
  - Sample clusters, then within selected clusters, sample people

Simple Random Sample

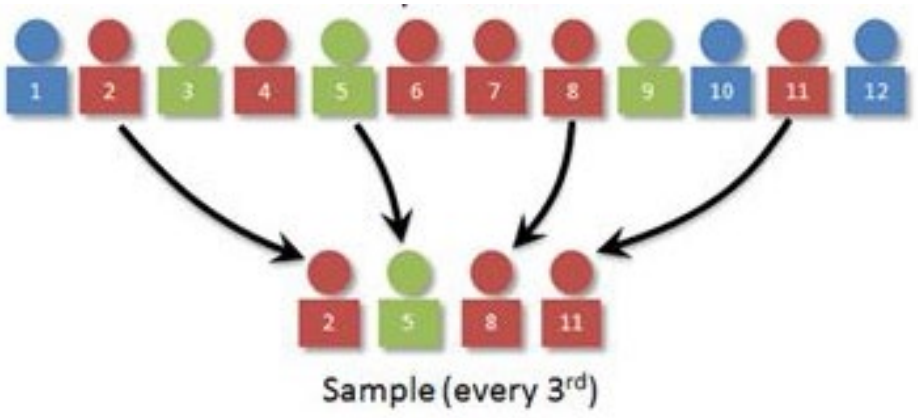
- If you have a comprehensive sampling frame
- Calculate the sampling fraction
  - I need 1,000 people to achieve appropriate power
  - There are 1,000,000 people in my target population (and therefore my sampling frame)
  - Each person has a 1,000/ 1,000,000= 1 per 1,000 chance of being selected
  - Each person in my sample represents 1,000 people in the population
- Implementation is easy
- Analysis is easy

Simple Random Sample: Disadvantages

- You will miss rare subgroups or not have enough people in those subgroups to analyze
  - Racial/ethnic minorities
- Inefficient (i.e., bigger standard errors) if there is substantial heterogeneity in the population
  - Suppose the population prevalence of my disease is actually 2%.
  - In my 1,000 person sample, on average, 20 cases.
  - What are the chances I will have <11 cases?

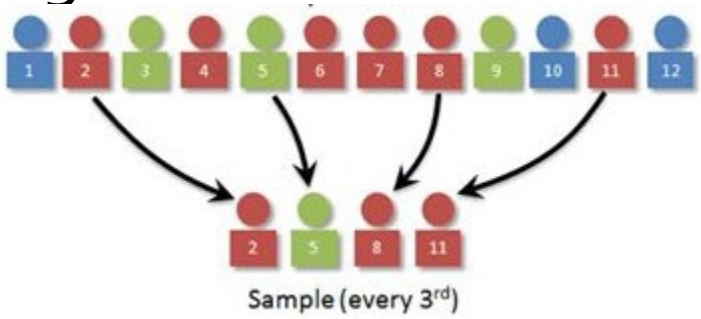
Systematic Sampling

- Calculate the sampling fraction you need (e.g., 1 /1000 or 1/3 or 1/k)
- Order the sampling frame list in some sequence
- Take every k^{th} person on the sampling frame list



Systematic Sampling

- Can be convenient if you are not certain of the list in advance
  - People attending clinic who screen in for food insecurity



  - Can go terribly awry if there's a pattern to the sequence that you did not anticipate, e.g., every 3rd patient sees Dr. V, who provides direct food pantry referrals

Stratified Sampling

- Within your sampling frame, identify a grouping variable, e.g., sex, race, age (something you know from the sampling frame or can learn before enrolling the person).
- Select a sampling fraction and choose that sampling fraction within each stratum
  - This ensures that if say your population is 50% men and 50% women, your sample will also be 50/50.
  - If your population is 70% white and 30% Latino, your sample will also be 70/30.
  - You eliminate the risk that, by chance, you might get 90% white and 10% Latino in your sample.
  - If you are trying to characterize something about the population which is very different for whites than Latinos, this improves efficiency.
- OR…Choose a different sampling fraction for some groups than others.

Stratified Sampling

- Within your sampling frame, identify a grouping variable, e.g., sex, race, age (something you know from the sampling frame or can learn before enrolling the person).
- Choose a different sampling fraction for some groups than others.
  - E.g., in a study of older adults:
    - 1/100 white women ages 65-74
    - 1/50 white women ages 75+
    - 1/75 white men ages 65-74
    - 1/25 white men ages 75+
    - 1/10 black women ages 65-74
    - 1/2 black women ages 75+
    - 1/5 black men ages 65-74
    - 1/1 black men ages 75+

Stratified Sampling

- Improves ability to study diversity and heterogeneity
  - If you only enroll 1/100 black men ages 75+, you may not have any in your sample.
- Improves efficiency
- Harder to implement
- Harder to analyze *if* unequal probability of selection
  - To describe the target population, you need to reweight by the inverse of the probability someone was selected.

1/100 white women ages 65-74 → each 1 in the sample represents 100 people in the popln

1/50 white women ages 75+ → each 1 in the sample represents 50 people in the popln

1/75 white men ages 65-74 → each 1 in the sample represents 75 people in the popln

1/10 black women ages 65-74 → each 1 in the sample represents 10 people in the popln

1/2 black women ages 75+ → each 1 in the sample represents 2 people in the popln

1/5 black men ages 65-74 → each 1 in the sample represents 5 people in the popln

1/1 black men ages 75+ → each 1 in the sample represents 1 person in the popln

Stratified Sampling

- Harder to analyze *if* unequal probability of selection
  - To describe the target population, *you need to reweight by the inverse of the probability someone was selected*.

1/100 white women ages 65-74 → each 1 in the sample represents 100 people in the popln

1/50 white women ages 75+ → each 1 in the sample represents 50 people in the popln

1/75 white men ages 65-74 → each 1 in the sample represents 75 people in the popln

1/10 black women ages 65-74 → each 1 in the sample represents 10 people in the popln

1/2 black women ages 75+ → each 1 in the sample represents 2 people in the popln

1/5 black men ages 65-74 → each 1 in the sample represents 5 people in the popln

1/1 black men ages 75+ → each 1 in the sample represents 1 person in the popln

- This concept will return to haunt us (or, empower us) as a tool to address study attrition and confounding using inverse probability weighting.

Horvitz-Thompson (1952) Estimator of a Population Total

$$\hat{Y} = \sum_{i=1}^{N} \frac{\delta_i Y_i}{p_i} = \sum_{i=1}^{n} \frac{y_i}{p_i}.$$

Where δ_i=1 if the element is included in the sample and 0 otherwise
And p_i is the probability of inclusion in the sample for element i.

1/100 white women ages 65-74 → p_i =.01
Woman 1, age 66: y=30
Woman 2, age 72: y=20

1/50 white women ages 75+ → p_i=.02
Woman 3, age 78: y=10
Woman 4, age 80: y=15

Estimated Total Y=(30/.01)+(20/.01)+(10/.02)+(15/.02)=6,250

Cluster or Multi-Stage Sampling

- Group the population into clusters
  - Clusters must be mutually exclusive (no person in the population can be in two clusters) and comprehensive (every person in the population must be in one cluster)
- Randomly select some of the clusters
  - Can use a simple random sample of clusters, or
  - Have different selection probabilities for some clusters
    - Select NY, CA, and TX with 100% chance
    - Select states with >40% rural population with 1/4 chance
    - Select all remaining states with 1/5 chance
- Within the selected cluster, list all eligible elements
  - Could also divide the cluster into comprehensive and mutually exclusive subclusters
- From the list of all eligible elements within the selected cluster, select a sample
  - Could be a simple random sample or a stratified sample

Cluster or Multi-Stage Sampling Advantages

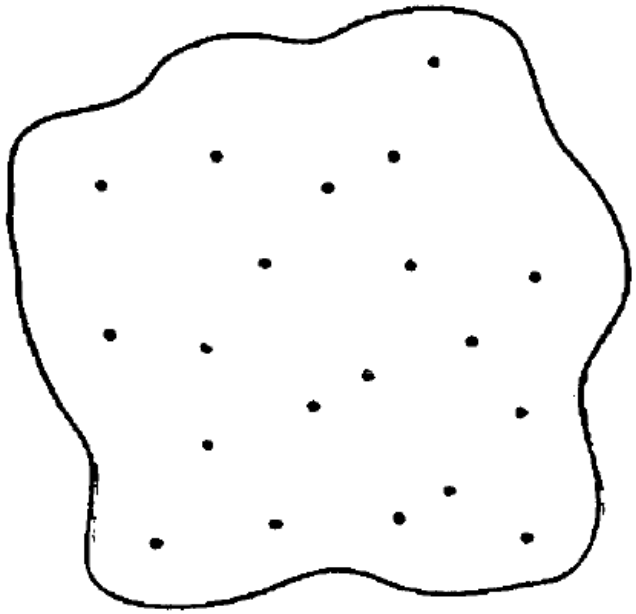
- You don't have to have a complete sampling frame in advance!
- Often much more logistically efficient
  - compare collecting a simple random sample of 1/100,000 people across the US vs
  - a clustered sample, where you first chose 20 states, then 5 counties in each state, then 300 people in each county
- Can incorporate stratification and unequal probabilities to learn more about subgroups

Cluster or Multi-Stage Sampling Disdvantages

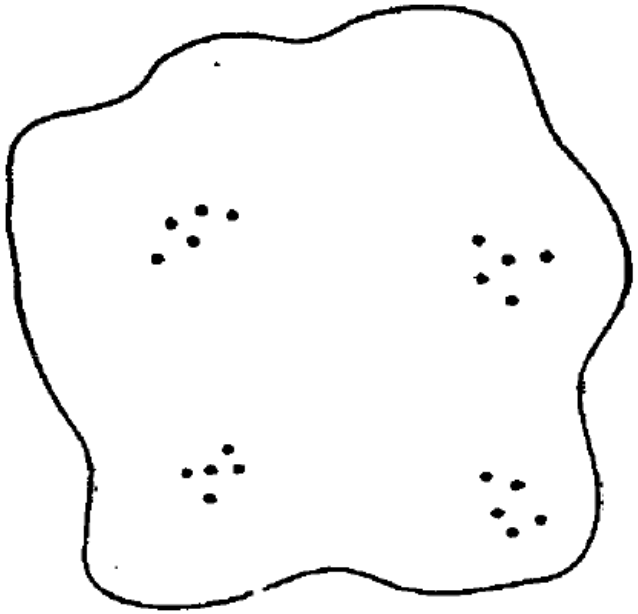
- You don't have to have a complete sampling frame in advance!
  - But you must have a comprehensive, mutually exclusive set of clusters
  - And be able to identify or create a sampling frame within each selected cluster
- Often much more logistically efficient
  - But statistically LESS efficient
  - Because people within clusters will tend to be more similar to one another than people in different clusters
  - So they are not independent observations.
  - If two people in a cluster are extremely similar, this is a big inefficiency
  - Usually not so similar though
- Have to deal with this mess in analysis to get the estimates and standard errors right

Sampling design creates clustering

a) Simple random sample



b) Two stage sample



Cluster or Multi-Stage Sampling Disdvantages

- Have to deal with this mess in analysis to get the estimates and standard errors right
- Sampling probability:
  - Probability your cluster was selected **x** probability you were selected, given that your cluster was selected. Suppose:
  - Chance your state was selected: 1/4
  - Chance your county was selected, given your state was selected: 5/72
  - Chance you wee selected, given your county was selected: 100/3000

Your chance of being selected = .25 * .069 * .0333=.000575

Your weight = 1/ .000575 = 1,739.

You represent 1,739 people.

Cluster or Multi-Stage Sampling Disdvantages

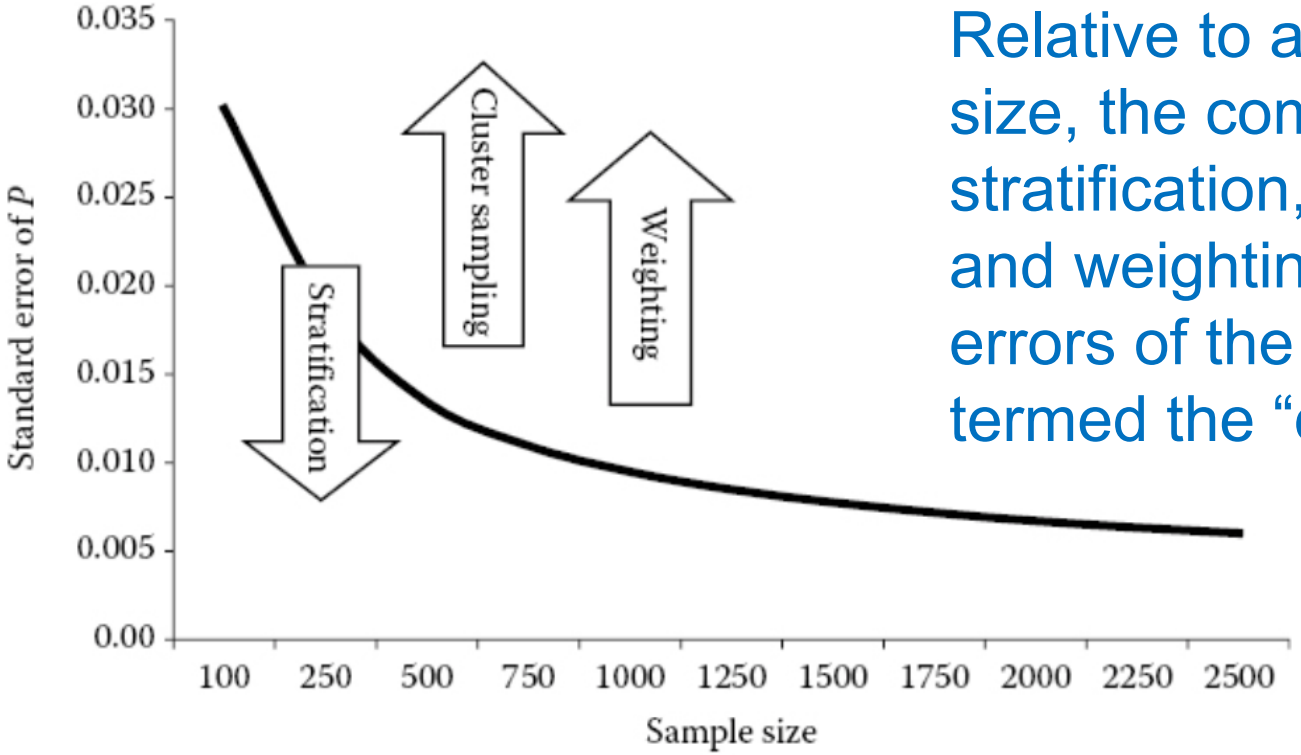
- Have to deal with this mess in analysis to get the estimates and standard errors right
- Sampling probability:
  - Probability your cluster was selected **x** probability you were selected, given that your cluster was selected. Suppose:
  - Chance your state was selected: 1
  - Chance your county was selected, given your state was selected: 10/50
  - Chance you wee selected, given your county was selected: 10/40

Your chance of being selected = 1 * .2 * .25=.05

Your weight = 1/ .05= 20.

You represent 20 people.

Complex Sampling Alters Variance of Estimates for Population Characteristics



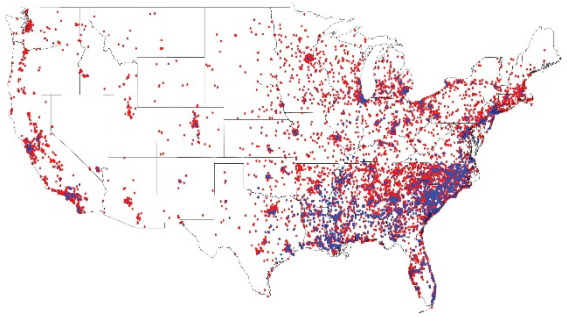
Relative to a sample of equal size, the complex effects of stratification, cluster sampling, and weighting on the standard errors of the estimates are termed the "design effects"".

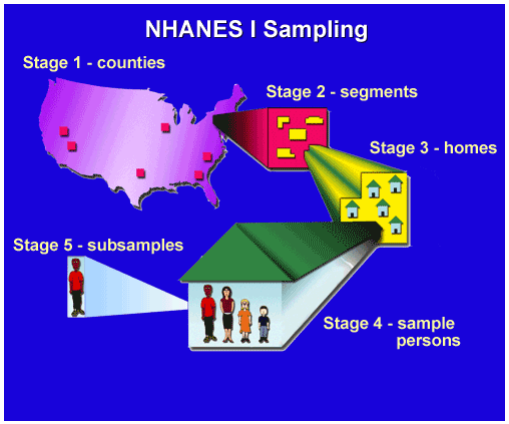
Figure 2.2
Complex sample design effects on standard errors of prevalence estimates.

From Heeringa, West, and Berglund, Applied Survey Data Analysis, 2nd Edn

Why use complex sampling?

- Nearly every major probability based sample uses either clustering or stratification or both.
- Stratified sampling essential to be able to study small(er) subgroups, e.g., old men.
- Cluster sampling less expensive if all of your participants live near each other= reduced survey time
- Cluster sample implies no need for census of the sampling frame: just need a census of the clusters





Steps in sampling plan: recap

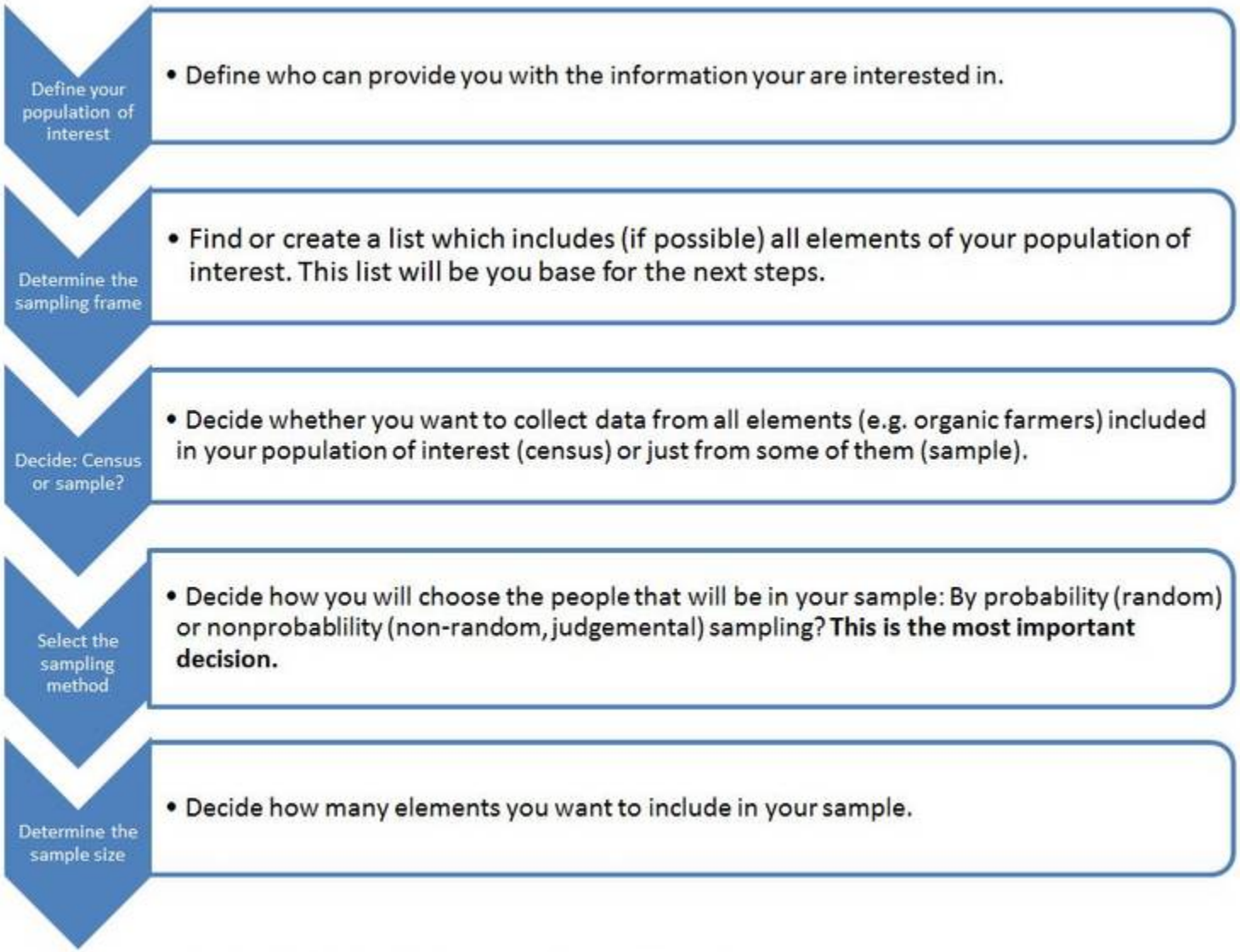


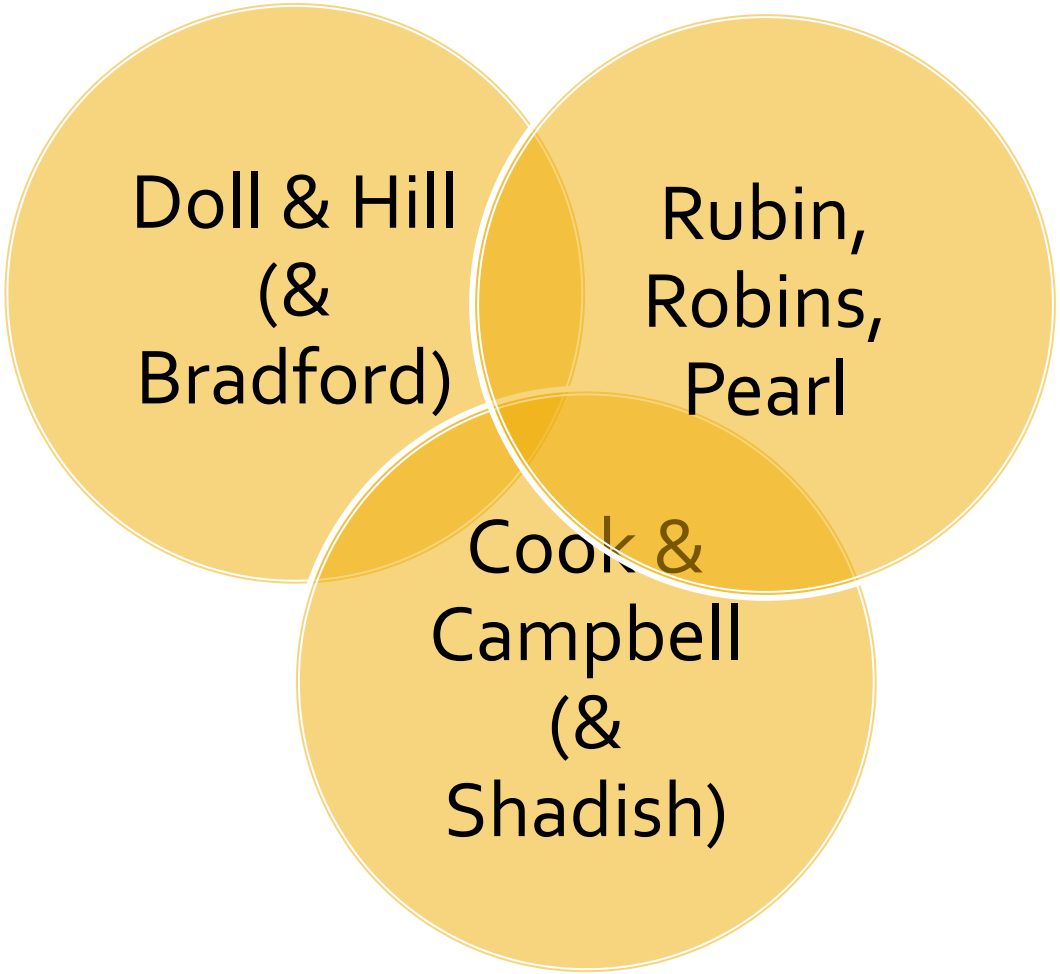
Figure 2: Overview on the sampling process, based on Malhotra et al. (2012)

Clustering and weighting are not just for sampling problems

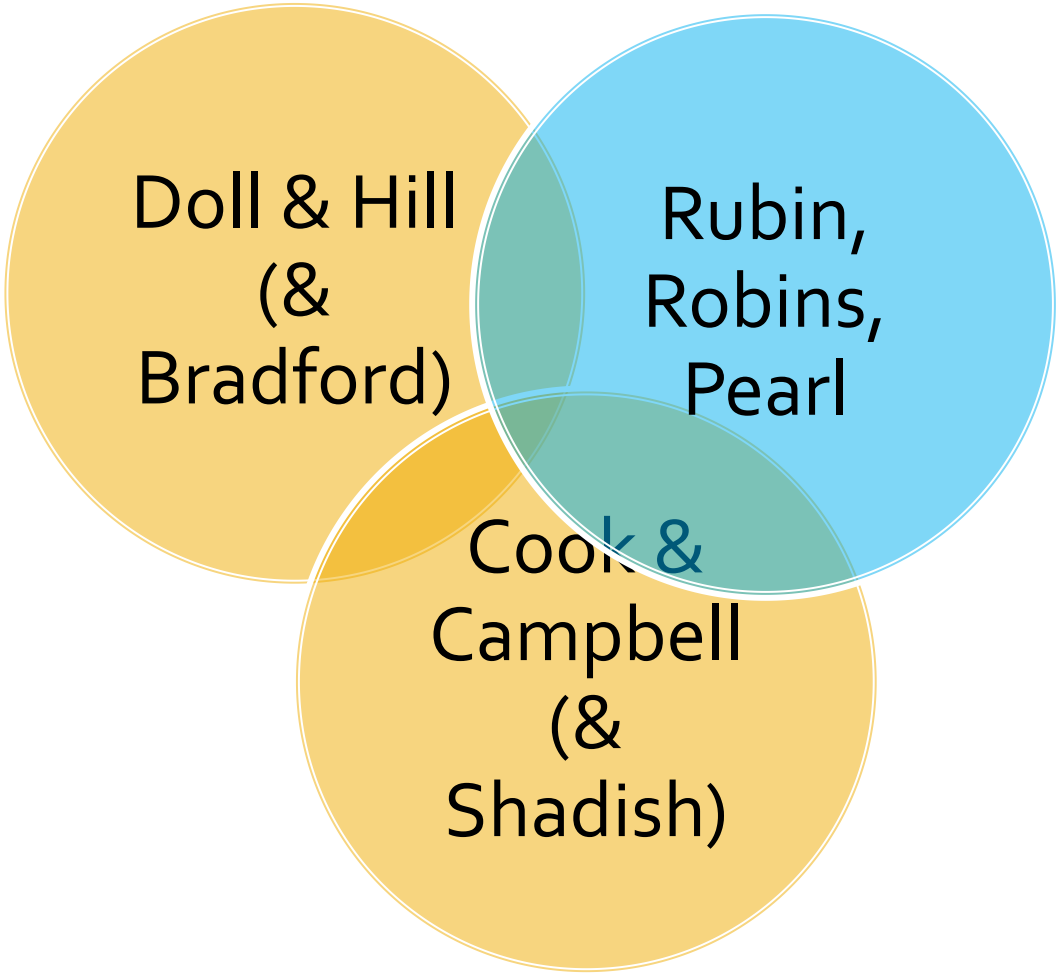
- Many other phenomena create clusters, including cluster randomization (next week), contexts where people live/are treated (neighborhoods, schools, clinics, hospitals), or repeated measures on the same person.
- In each case how to deal with the correlation between observations – whether it is a nuisance or substantively interesting – is a key decision.

Comparing Causal Inference Frameworks

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)

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

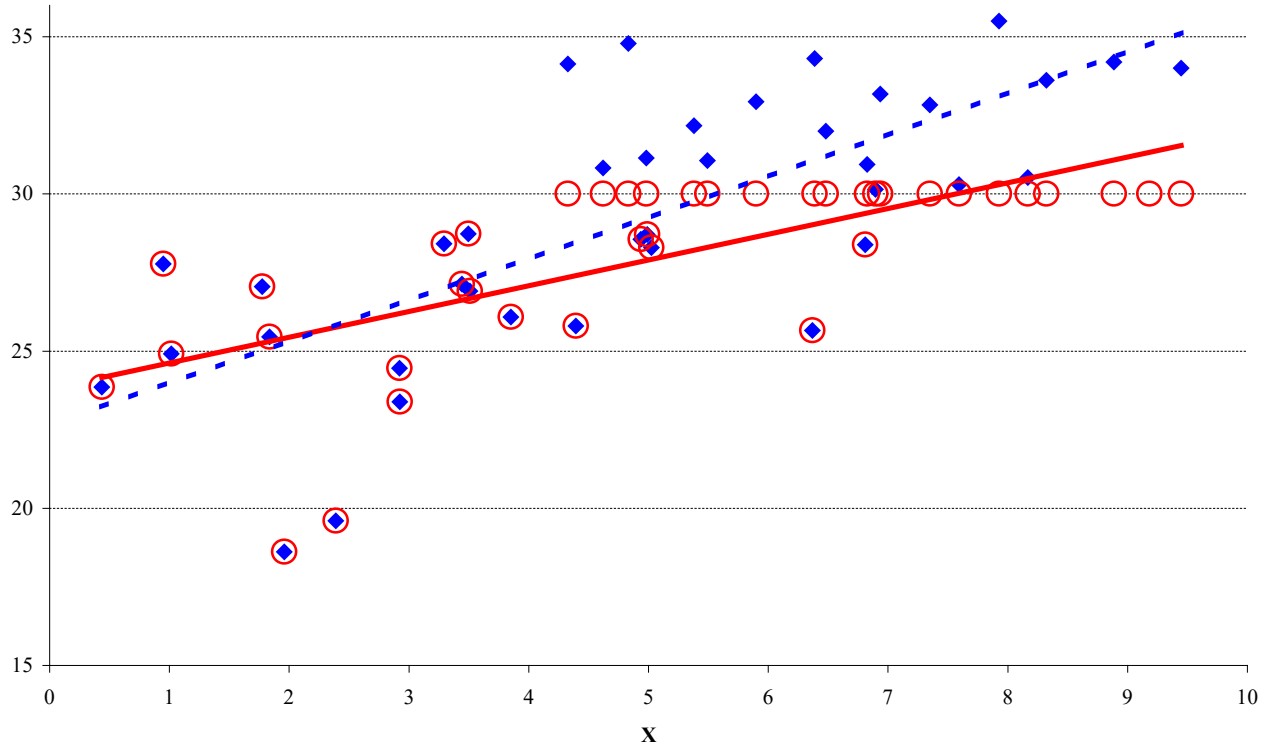
Threats to Statistical Conclusion Validity: Examples and opportunities to address?

- 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?

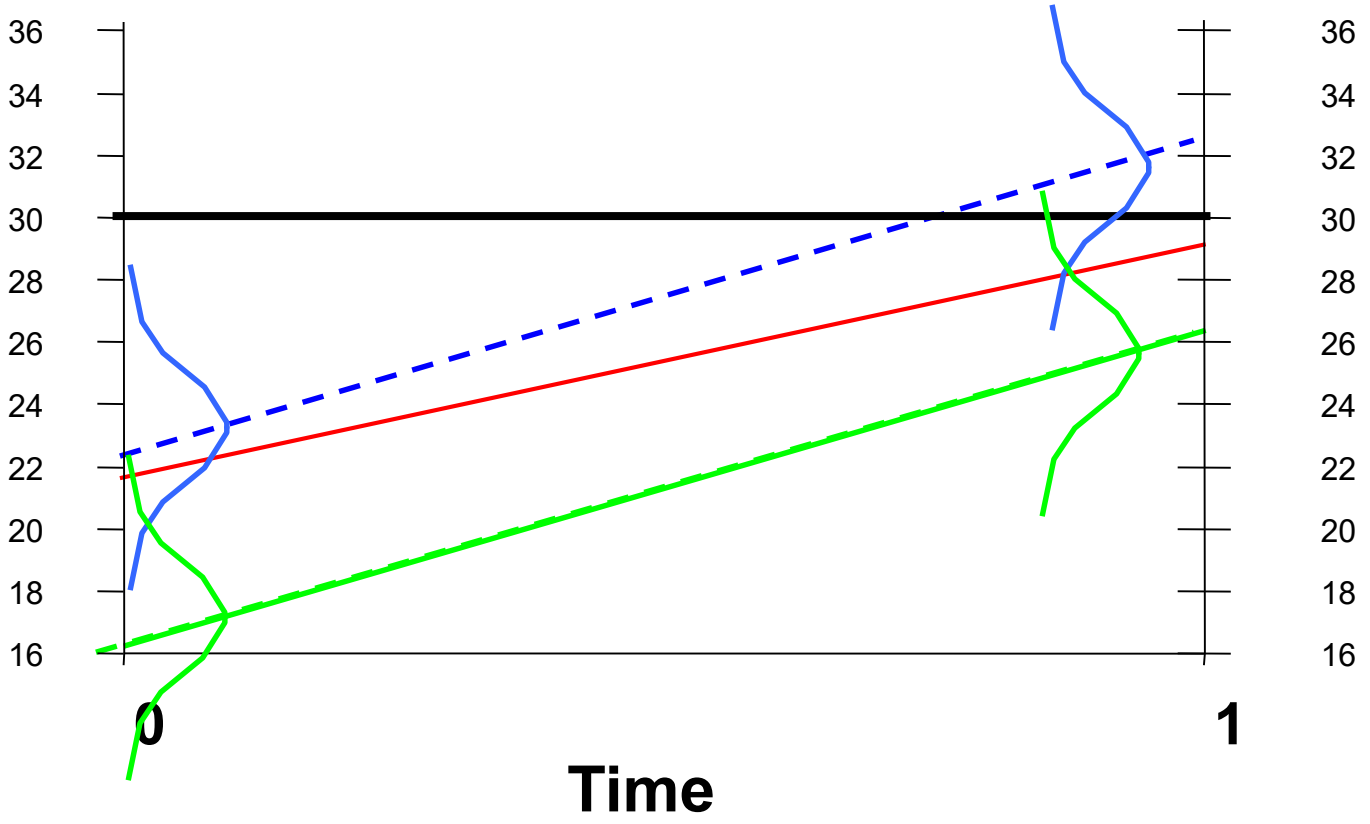
- Type 1 error rate: probability of seeing an association as extreme as this under the null
- Type 2 error rate: probability of failing to detect an association of specified magnitude with given design (sometimes called β)
- Sample size: number of *independent* observations
  - In a clustered sample: 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 Data: 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?

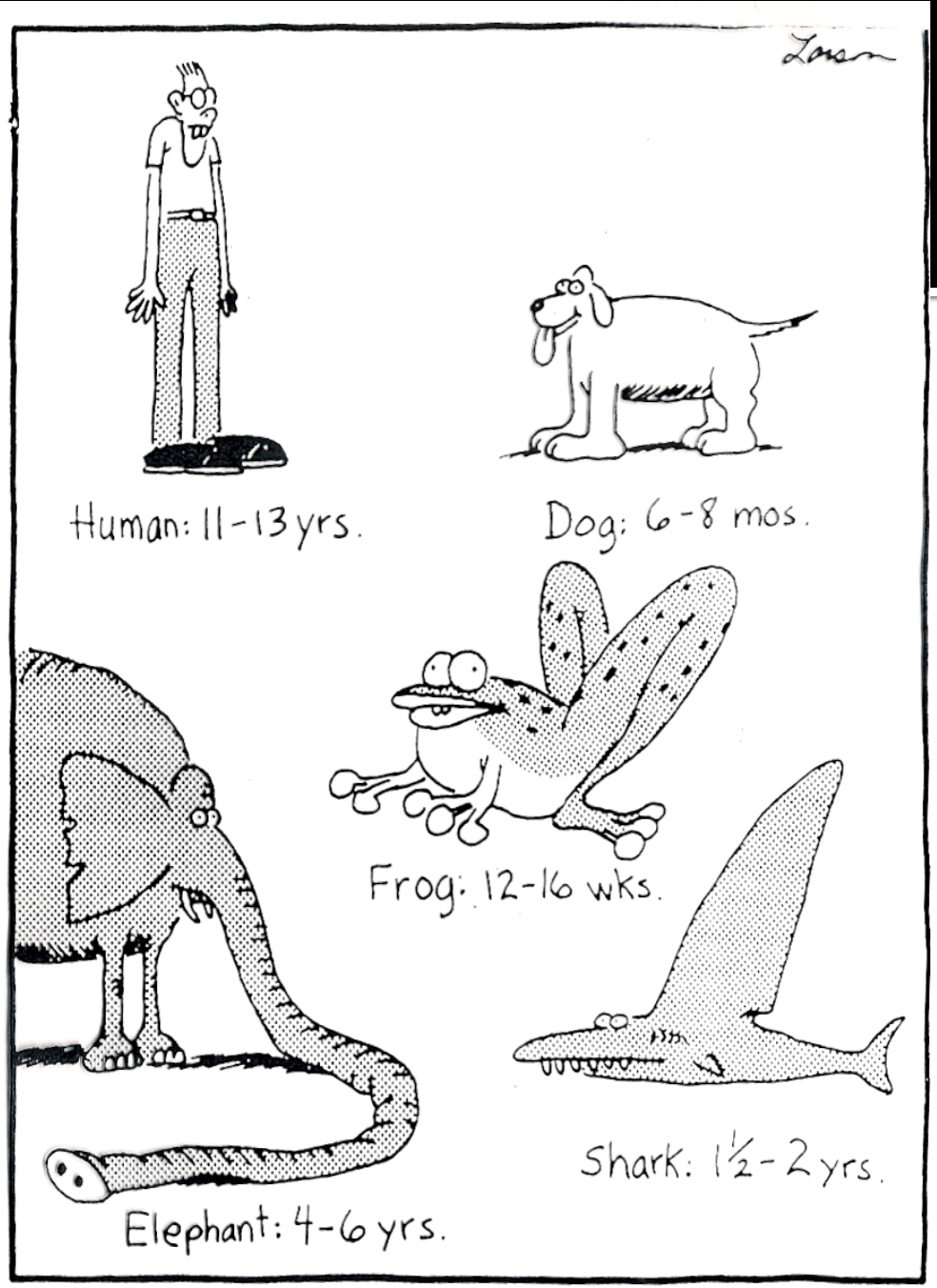
Threats to Internal Validity

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



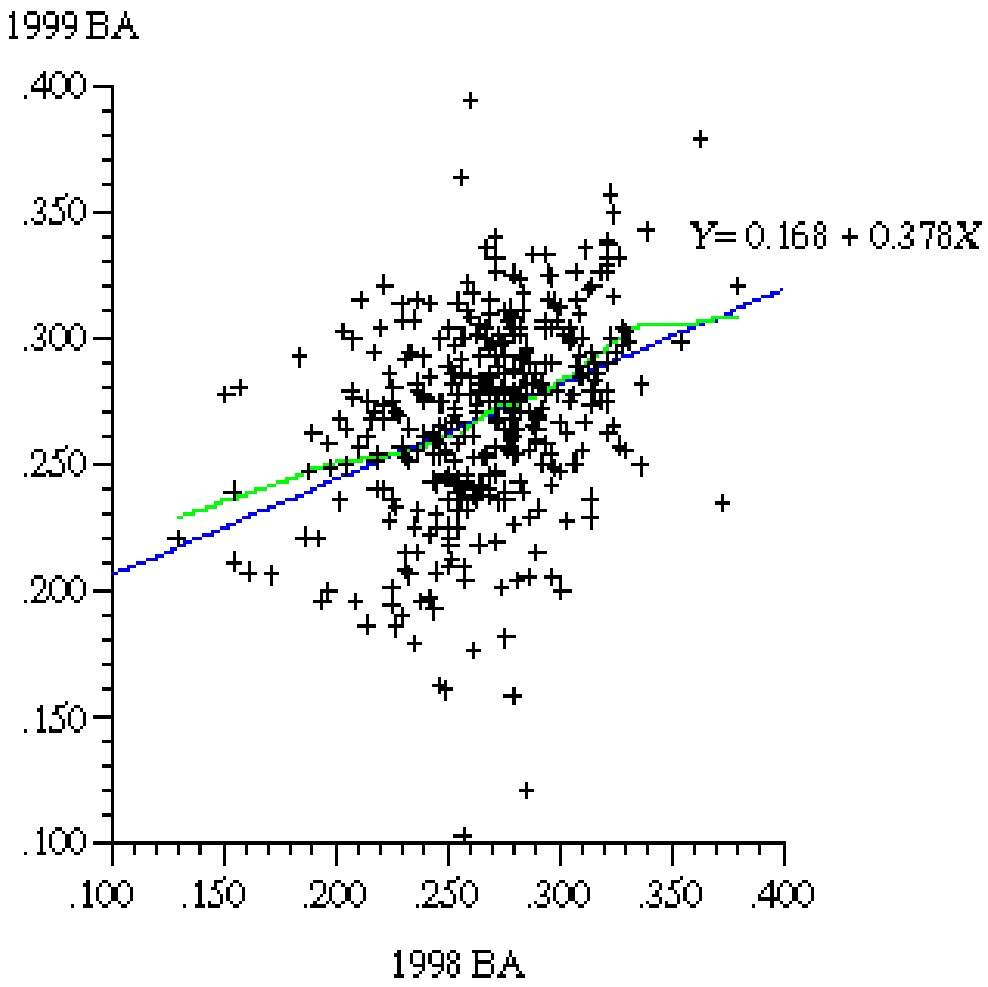
Awkward ages

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.

Types of Validity: Construct Validity

3. Construct validity: the validity of inferences about the higher order constructs that you thought you were measuring
   - Definition & measurement of exposure and outcome?
   - Psychometric methods for establishing validity and reliability and correcting for lack thereof
   - We will not spend much time on technical tools to address these issues in this class, but it is important to recognize potential issues.

Types of Validity: External Validity

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.
   - Generalizability, transportability
   - If lack of validity driven by heterogeneity in causal effects
     - Address this via reweighting from sample where you've assessed effects to the sample where you want to implement the intervention
     - Sample with a priori known probability of inclusion for each person in the population (e.g., NHANES) and use this probability to reweight
     - Estimate the probability of inclusion post-hoc and use this probability to reweight
   - Lack of external validity may also reflect differences in implementation (gets fuzzy distinguishing this from construct validity).

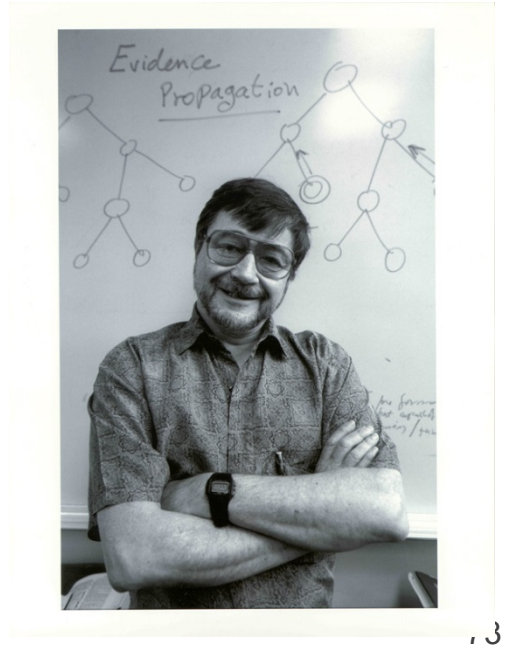
"Causal Inference Methods" in epidemiology dominated by Robins (occ health MD), Pearl (computer scientist) and trainees

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.



Counterfactual/modern epi framework emphasizes 3 assumptions for causal inference

- Exchangeability: exposure each individual receives is unrelated to that person's potential outcome if s/he receives any particular exposure
  - Violated if a confounder influences which exposure you receive and also influences your outcome
- Consistency: Your potential outcome under any particular treatment equals your actual outcome if you received that treatment
  - Violated if there are multiple "flavors" of treatment, all labeled the same thing "the treatment" and your outcome depends on which flavor you receive
- Positivity: everyone has a probability greater than 0 but less than 1 of receiving each version of treatment
  - Violated if some types of people are certain to be treated because then I have nobody to compare them to, and cannot hope to estimate what their outcomes would have been if they hadn't been treated.

Counterfactual/modern epi framework emphasizes 3 assumptions for causal inference

- Exchangeability: exposure each individual receives is unrelated to that person's potential outcome if s/he receives any particular exposure
  - Violated if a confounder influences which exposure you receive and also influences your outcome
- Consistency: Your potential outcome under any particular treatment equals your actual outcome if you received that treatment
  - Violated if there are multiple "flavors" of treatment, all labeled the same thing "the treatment" and your outcome depends on which flavor you receive
- Positivity: everyone has a probability greater than 0 but less than 1 of receiving each version of treatment
  - Violated if some types of people are certain to be treated because then I have nobody to compare them to, and cannot hope to estimate what their outcomes would have been if they hadn't been treated.
- **Bonus assumption: Correctly specified model!**

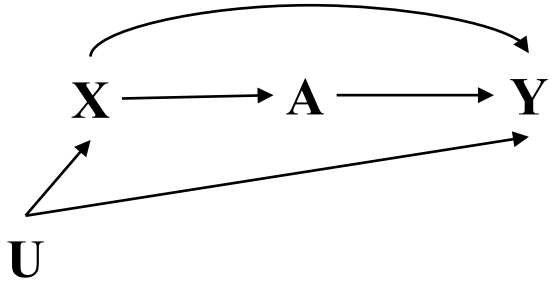
Counterfactual/modern epi framework emphasizes 3 assumptions for causal inference

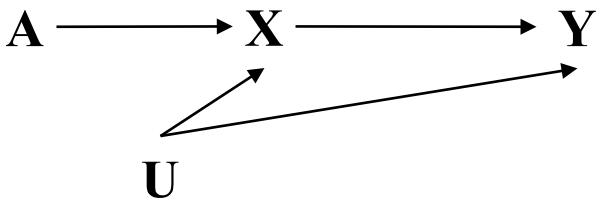
- Exchangeability: exposure each individual receives is unrelated to that person's potential outcome if s/he receives any particular exposure
  - Violated if a confounder influences which exposure you receive and also influences your outcome
- Consistency: Your potential outcome under any particular treatment equals your actual outcome if you received that treatment
  - Violated if there are multiple "flavors" of treatment, all labeled the same thing "the treatment" and your outcome depends on which flavor you receive
- Positivity: everyone has a probability greater than 0 but less than 1 of receiving each version of treatment
  - Violated if some types of people are certain to be treated because then I have nobody to compare them to, and cannot hope to estimate what their outcomes would have been if they hadn't been treated.
- **Bonus assumption: Correctly specified model!**
- **Hey one more! Assume an infinite sample size.**

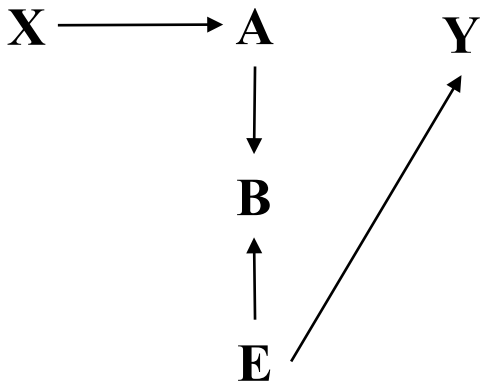
Causal Directed Acyclic Graphs are a useful tool to evaluate internal validity given your prior assumptions.

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

•We are not going to review DAGs in this class, but if you are not confident, read ME3 chapter, or chapter in social epi methods.







Where do DAGs fit in Cook & Campbell "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. Causal inference framework avoids getting into this with our bonus assumptions:

- **Bonus assumption: Correctly specified model!**
- **Hey one more! Assume an infinite sample size.**

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 and measurement 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?

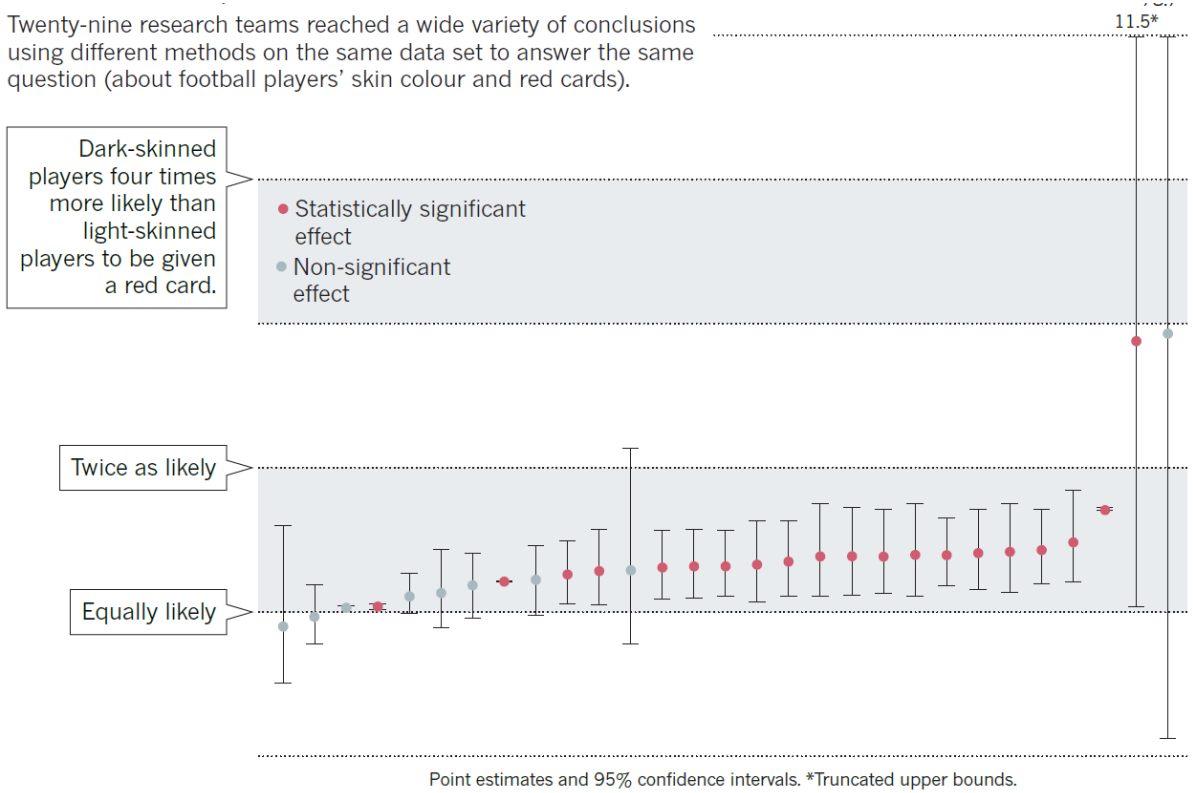
These are all *helpful* to think about, but recognize that only temporality is really necessary. All others are to help make an informed judgment, but they are not "criteria":

- 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.

End Lecture 1