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

Greatest hits. Most confusing points. What I hope you remember.

- Plus….

Jean's questions

- Can you explain slide 44 of uncertainty lecture on relationship between CIs and p-values?
- What is the difference between sufficient-cause interaction and epistatic interaction?
- Why for natural direct and indirect effect do we always set M to what if would be for the reference value of X (X = 0) rather than some other value of X (e.g. X = 1)?
- Can you review why we would want the controlled vs. natural effects in mediation?
- Is this correct: If there is no effect modification between the indirect pathway and the direct pathway, then there is only one CDE and it is equal to the NDE?

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

What Kinds of Question Do You Have?

- **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?

If you have causal questions (which hopefully you do)

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

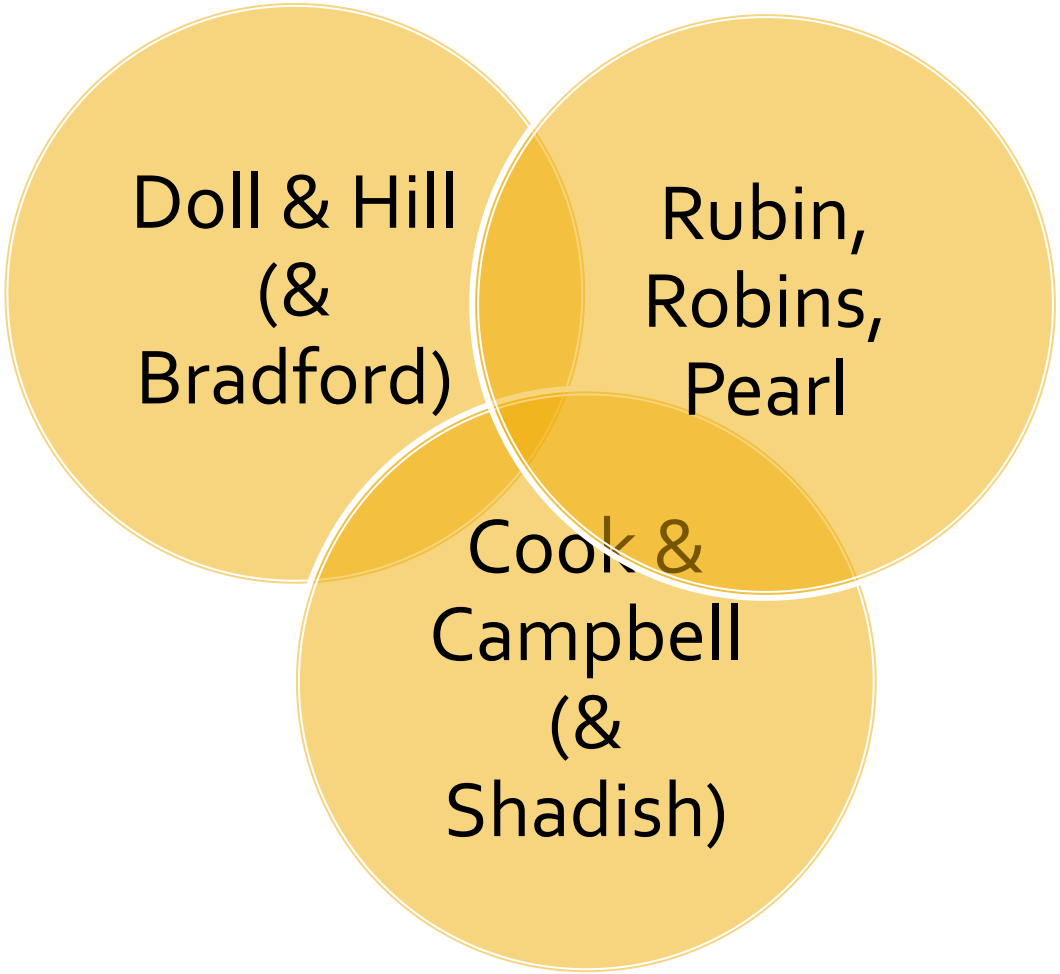
Think about the causal roadmap (from Petersen and van der Laan 2015 Roadmap)

- Specify what you know about the causal system
  - Use a DAG, structural causal model, or both
- Specify the data you observed and how you got them: how were the observations selected into your data set and what was measured?
- **Specify the target causal quantity**
  - **A counterfactual contrast and causal estimand**
- Can your causal quantity be identified? (eg Did you fulfill the back door criterion? The front door criterion? The IV assumptions?) If not, or you're not sure, what assumptions would you need to make to identify the estimand (eg, no confounders).
- Describe how to estimate the causal quantity with a statistical parameter
- Implement the analysis you proposed
- Interpret findings, clarifying the causal and statistical assumptions under which your estimate corresponds to a causal parameter.
  - Note which assumptions are shaky

Specify the target causal quantity A counterfactual contrast and causal estimand (from Petersen and van der Laan 2015 Roadmap)

- What will you intervene on?
  - Intervene once or repeatedly?
- How will you intervene?
  - Set the variable to a particular value?
  - Follow a rule for setting the value, which may result in different people having different values?
  - Choose a distribution of the variable across the population?
- How will you characterize the contrast of counterfactuals?
  - A ratio of the average value under two treatment regimes?
  - A difference in the average value under two treatment regimes?
  - Conditional effects (eg each person versus him or herself) or marginal effects (eg whole population vs itself)
- Who will you intervene on? What is the target population?
  - Everyone?
  - Subset, e.g., the compliers?
  - Different people, not entirely like the people in your sample?

Alternative traditions for discussing causality



How do these three traditions relate?

- Rubin/Robins/Pearl: express contrasts as counterfactuals (or similar), formalize the link between causation and association, describe formal assumptions necessary to draw causal inferences.
- Cook, Campbell, & Shadish: No formalisms. Practical experience and clear thinking about what can go wrong when trying to draw causal inferences without a perfect RCT. Separates into statistical, construct (measurement), internal, and external validity problems.
- Doll & Hill: for research consumers to worry about drawing inferences based on a body of literature or a specific study.

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?

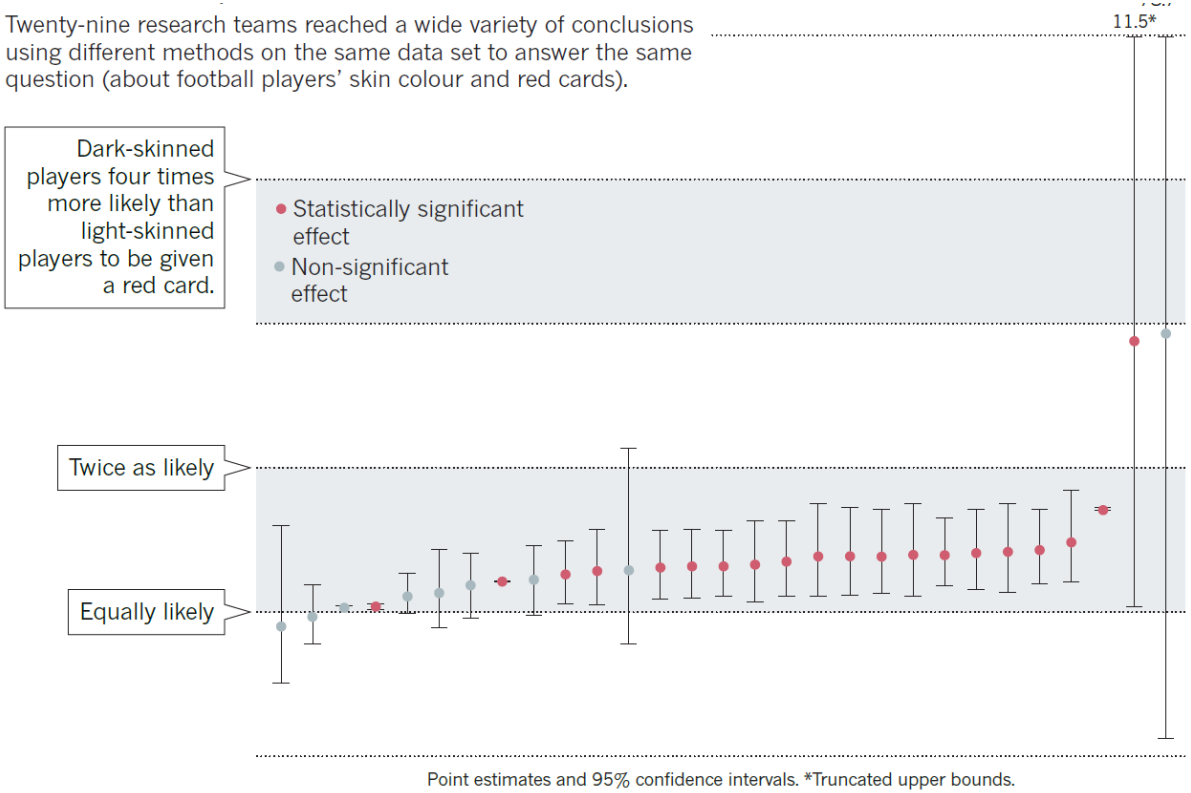
These are all *helpful* to think about, but recognize that only temporality is really necessary. All others are to help make an informed judgment, but they are not "criteria":

- 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.

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).

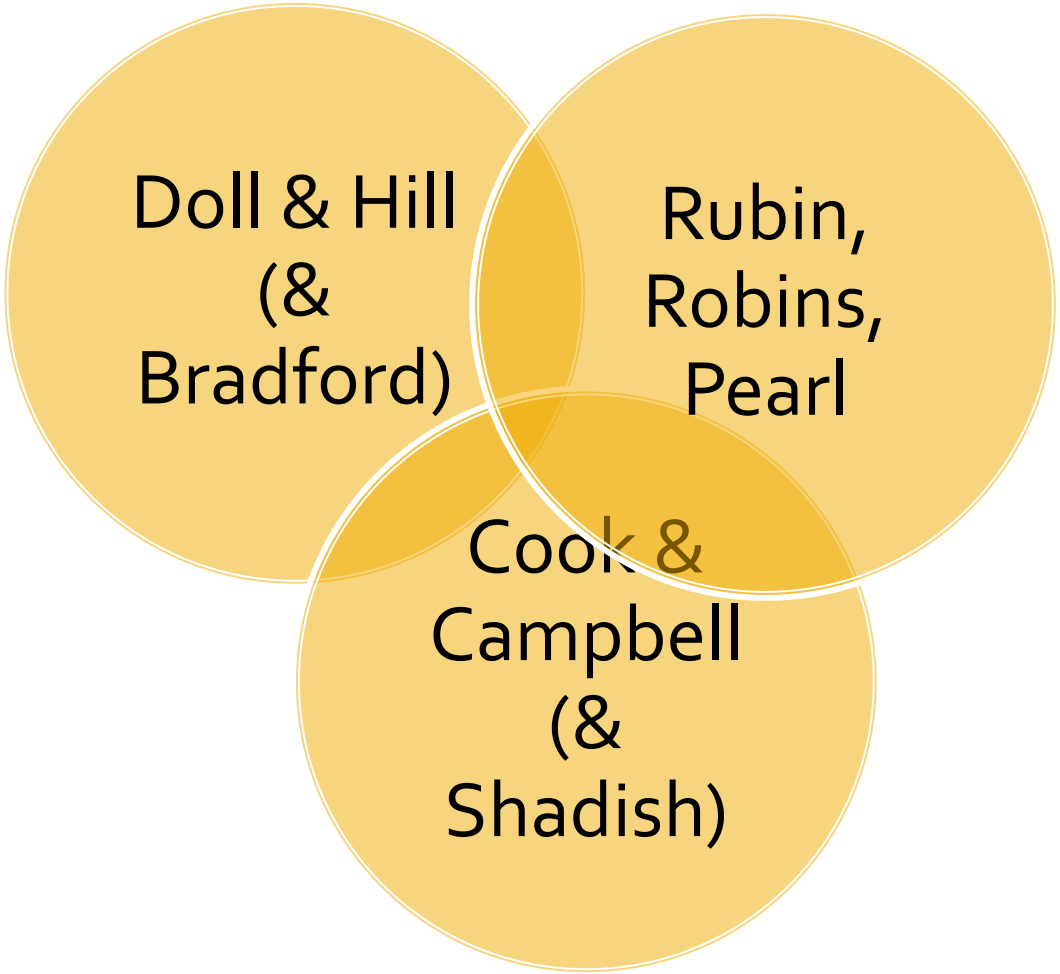
Counterfactual/modern epi framework emphasizes 3 assumptions for causal inference

- 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.**

How do these three traditions relate?

- Rubin/Robins/Pearl: express contrasts as counterfactuals (or similar), formalize the link between causation and association, describe formal assumptions necessary to draw causal inferences. Focuses on Cook & Campbell's internal validity, with recent forays into external validity. Assume statistical conclusion validity (someone else's problem). Assume you know how to draw the DAG.
- Cook, Campbell, & Shadish: No formalisms. Practical experience and clear thinking about what can go wrong when trying to draw causal inferences without a perfect RCT. Separates into statistical, construct (measurement), internal, and external validity problems. No use of DAGs (though you could draw these, see Matthay) and no acknowledgment of eg collider bias.
- Doll & Hill: for research consumers to worry about drawing inferences based on a body of literature or a specific study. Not actually right, but sometimes useful.

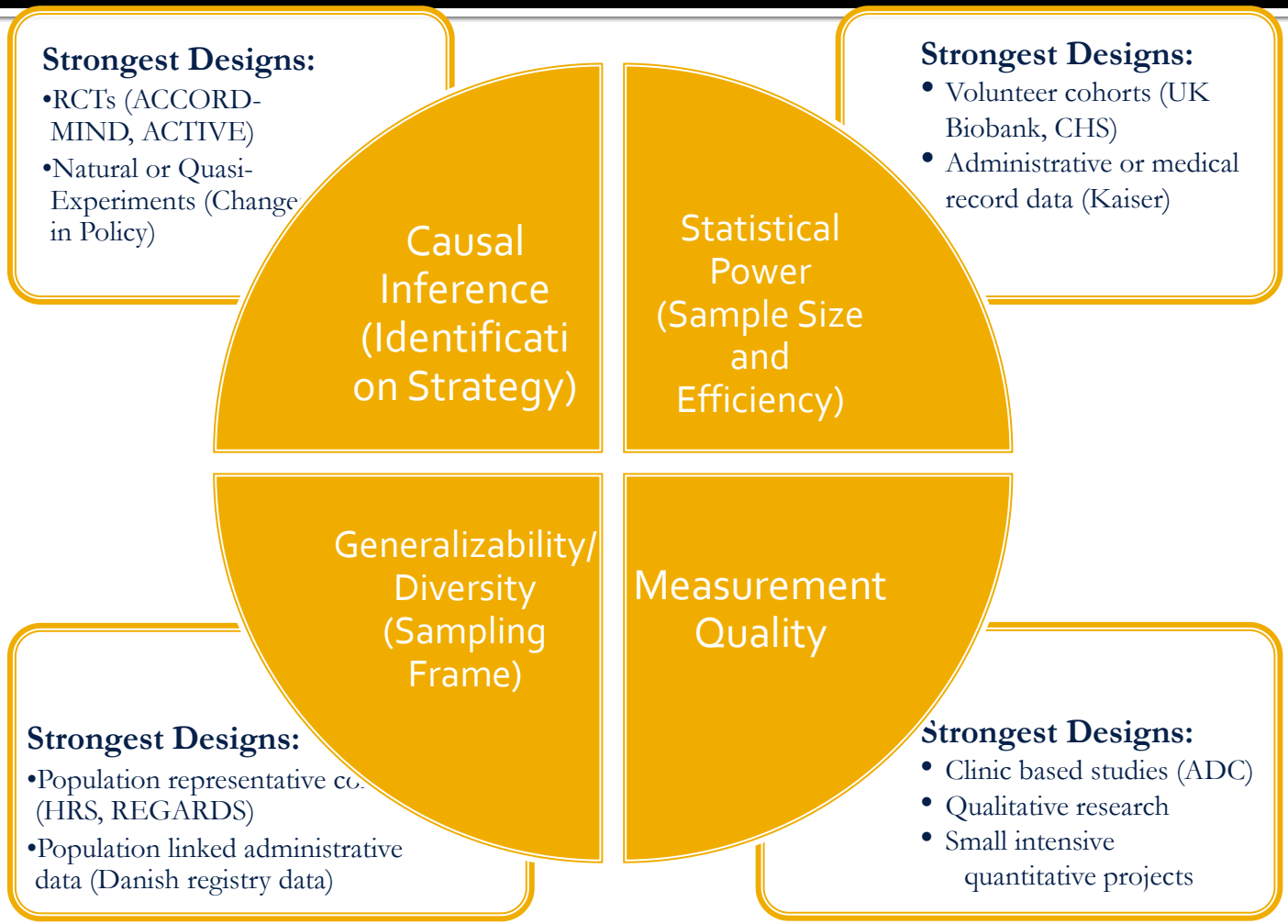
Alternative traditions for discussion causality



Data source tradeoffs

- Who is in the sample?
  - Lack of selection bias (predictors of outcome do not predict participation)
  - Diversity/representativeness of population of interest
- How many people are in the sample?
- What do they measure on those people?
  - Substantive relevance of measures
  - Quality of measures
  - Number of measures
- How does the sample limit or support the identification strategy?
  - Usually most important: longitudinal data

Optimal designs for these goals typically in tension with one another: Better measurement quality increases expense, reduces sample size.



Sampling Intro

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

Non-representative sampling could be fine but remember it can create entirely spurious associations and intractable confounding

The overlap of race and place for older Americans

		African-Americans		Whites	
		Birthplace		Birthplace	
		Non-Southern	Southern	Non-Southern	Southern
Current Residence	Non-Southern	15%	29%	60%	5%
	Southern	2%	54%	10%	25%
		17%	83%	70%	30%

2000 Census 5% sample, age 65+, race black or white, southern= census region.

Representative sampling

- Step 1: Who is the target population?
- Step 2: Do you need a census, a representative sample, or just some people from the target population?
- Step 3: Can you identify a sampling frame? If not, can you divide the target into comprehensive, mutually exclusive groups which you are able to list and then use for cluster sampling?
- Step 4: Choose the sample design

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

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

- 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

Count this person 100 times in the analysis.

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

Count this person 5 times in the analysis.

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

Count this person 1 time in the analysis.

- This is the basis of inverse probability weighting, which we use to analyze stratified samples (IPW to make the probability someone was selected independent of their other characteristics)
- The greater the variance of the weights, the worse the design effect, because the most efficient way to use your data is if each person counts their full selves.
- Roughly, the design effect when applying weights =
  - 1+(the variance of the weights)/(square of the mean of the weights)

Inverse probability weighting: not just for representative surveys!

- Use IPW to reweight a sample back to the population from which it was sampled, when you have a known sampling fraction
- If you don't know the sampling fraction, estimate it using known characteristics:

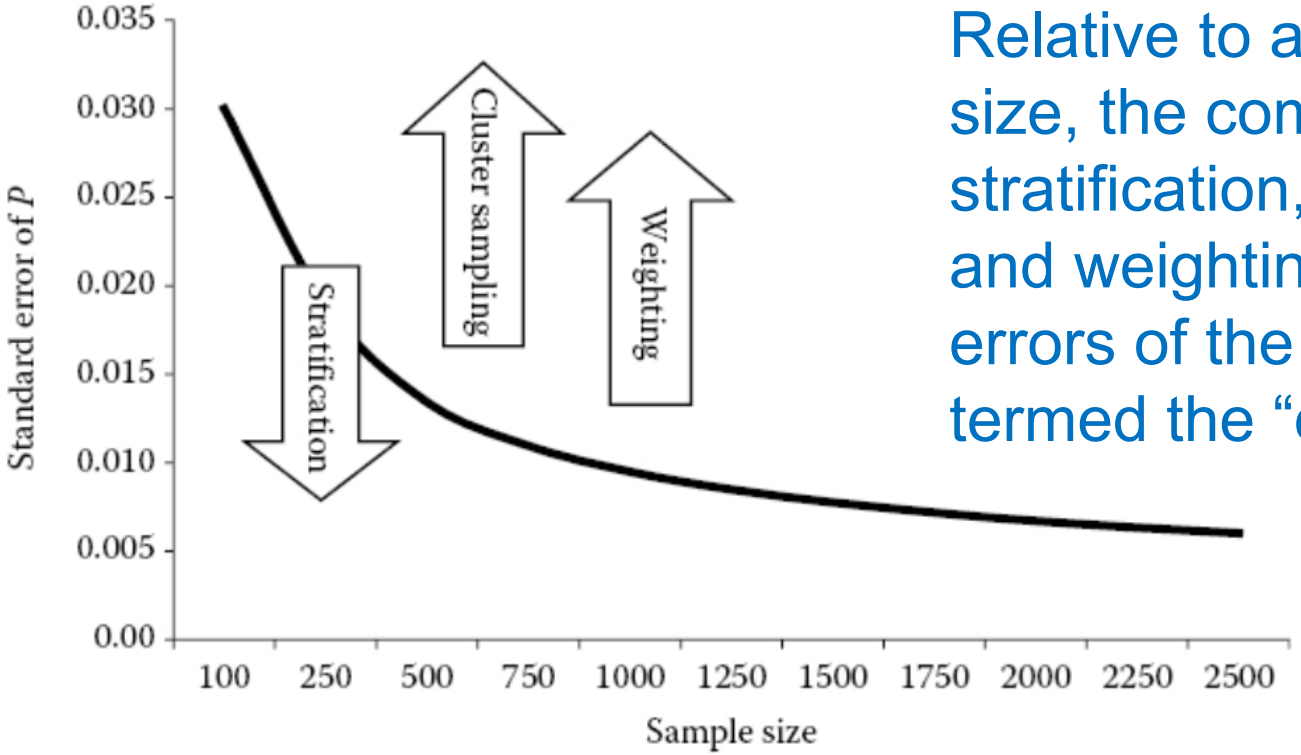
Pr(person was in my analytic sample | demographics of the person)=bo+b1*demographic_1+b2*demographic_2 etc.

- You can estimate this with logistic regression for example, if you know the demographic characteristics of the population
- You can apply the same idea to 'transport' an effect from your analytic sample to some other target population, even if your analytic sample was not drawn from that target.
- We also use IPW to control confounding (IPW to make the probability someone was treated independent of their other characteristics) and
- We use IPW as one way to correct selective attrition or missing data (IPW to make the probability someone attrited from the sample independent of their other characteristics).
- Same issue with the design effect/ impact on variance applies.

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

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".

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

Figure 2.2

Complex sample design effects on standard errors of prevalence estimates.

Examples of sources of clustering

- Contextual determinants of exposure or outcome: cluster randomized trials, neighborhood effects, school effects, families, clinics
- Sampling in places: neighborhoods, schools, hospitals, clinics, families, siblings
- Repeated measures over time on the same person
- Two eyes in the same person
- Multiple measures of related constructs in the same person (math scores, verbal scores)

Features of clustering

- Clusters may have equal numbers of observations (balanced) or unequal
- Items within a cluster may be ordered with respect to one another (e.g., time creates an ordering)
- Assume that conditional on cluster, observations are independent

Clusters can be a statistical nuisance or a scientific feature

- Statistical nuisance because the classic formulas to generate confidence intervals, p-values, or other measures of uncertainty assume all observations are independent.
- If the observations are correlated, generally you can extract less information than with independent observations.
- Clusters may also represent interesting substantive questions about the determinants of disease.
- Useful to think about this when choosing an analysis plan.

The nuisance part: design effect and effective sample size

- Standard sample size approaches leads to an underpowered study when observations are clustered

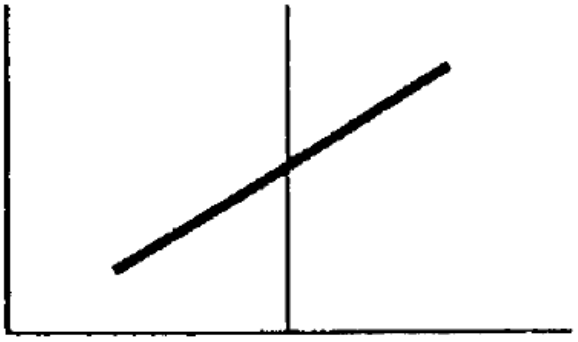
Effective Sample Size = km/DEff,
Where DEff = 1+ICC*(m-1)

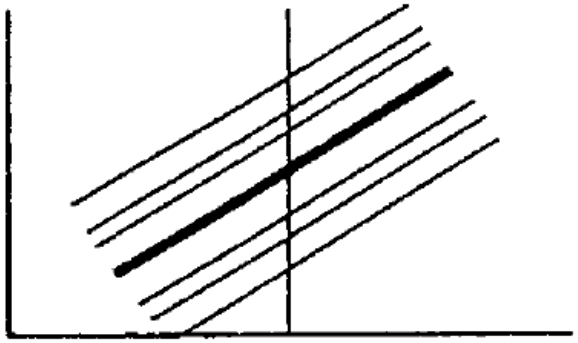
k=number of clusters
m=population per cluster

The nuisance part: plenty of options to fix the standard errors

- Mixed models (aka random effects models)
  - Allow you to ask new questions about the structure of the association between observations, which is often of substantive interest.
  - Handle unbalanced data very nicely
- GEE
  - Are consistent even if you are wrong about how the observations are correlated
- Robust variance calculations
- Because adding observations is expensive, when planning study designs calculate the impact of more clusters with fewer individuals or fewer clusters with more individuals.

Basic mixed model for data clustered within places (i.e., no ordering)





We assume places differ by some little bit (which we will call the mu for that place, or μ_{0j}) and everyone in the same place has this same little bit of offset, plus their own personal differences.

When predicting each person's outcome value, we add this place-specific little bit to the intercept, along with differences predicted by any other factor.

Basic mixed model for clustered data without ordering

Two level random intercepts model (i indexes individual, j indexes place; if there is an i subscript, it means this variable or coefficient differs between people; if there is a j in the subscript, it means this variable or coefficient varies between places. If there's a j but no i, it means this variable or coefficient has the same value for every person in the same place. :

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij} \text{ [Micro model]}$$

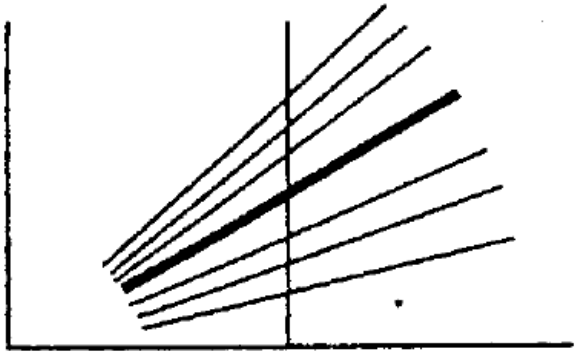
$$\text{[Overall model]} = y_{ij} = \beta_0 + \beta_1 x_{ij} + (\mu_{0j} + \varepsilon_{ij})$$

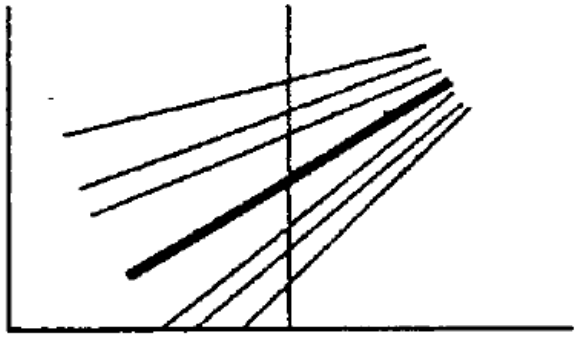
To fully specify a mixed model, you must also describe the distribution of the random terms

$\mu_{0j} \sim N(0, \sigma^2_{\mu 0})$

$\varepsilon_{ij} \sim N(0, \sigma^2_e)$

Random slopes model





Random slopes model (implicitly, assume you have a random intercept as well):

Not only do places differ by a little bit (the μ_{0j} random intercept) to the same extent for everyone in the place, but the effect of some exposure differs depending on the place (μ_{1j}), so the "slope", or coefficient is different.

Random slopes model

Two level random slopes model (implicitly, assume you have a random intercept as well):

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \varepsilon_{ij} \text{ [Micro model]}$$

$$\beta_{0j} = \beta_0 + \mu_{0j} \text{ [Macro model for intercept]}$$

$$\beta_{1j} = \beta_1 + \mu_{1j} \text{ [Macro model for slope]}$$

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (\mu_{1j}x_{1ij} + \mu_{0j}x_{0ij} + \varepsilon_{ij}x_{0ij})$$

Why do we prefer longitudinal data?

- Substantive question entails longitudinal data
  - Long etiologic period between exposure and outcome
  - Disease incidence or disease recovery processes are substantive outcomes
- To strengthen causal inference
  - Establish temporal order
  - Help to avoid adjusting for mediating variables or variables downstream of outcome
  - Address time-constant confounders (even ones that you did not measure)
  - None of these is a slam dunk

Longitudinal data help address time-constant confounders (even ones that you did not measure)

- Consider instead the "true" structural equation for changes in Y:

$$Y_{i,t}-Y_{i,t-1}=(\beta_0+\beta_1*X_{i,t}+\beta_2*C_{i,t}+\beta_3*U_i+\varepsilon_{Yi,t})-(\beta_0+\beta_1*X_{i,t-1}+\beta_2*C_{i,t-1}+\beta_3*U_i+\varepsilon_{Yi,t-1})$$

$$Y_{i,t}-Y_{i,t-1}=\beta_1*(X_{i,t}-X_{i,t-1})+\beta_2*(C_{i,t}-C_{i,t-1})+\beta_3*(U_i-U_i)+\varepsilon_{Yi,t}-\varepsilon_{Yi,t-1}$$

- The omitted time-constant confounder does not predict change in Y.
- Key assumptions:
  - The confounder does not change over time
  - The effect of the counfounder (β_3) does not change over time
  - (Equivalent): the confounder does not influence change in Y

- Given these assumptions, you can use longitudinal data to eliminate bias from confounders you did not even measure.

This is one of three ways we learned that you can identify a causal effect even with unmeasured confounding.

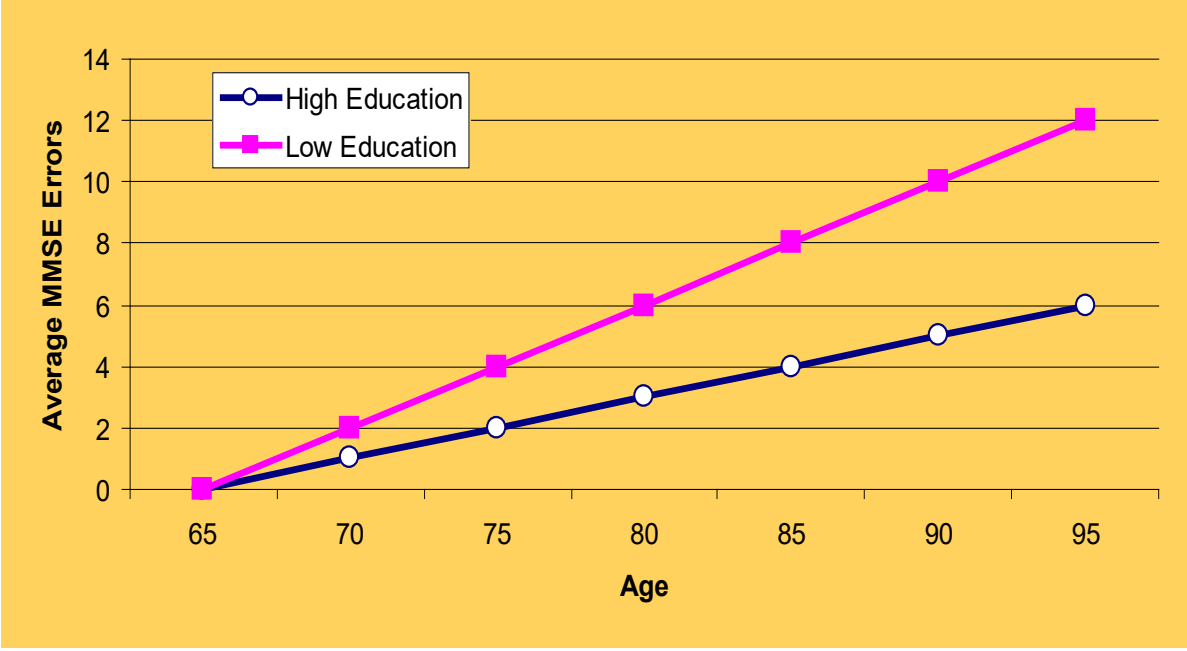
- The other approaches are
  - Instrumental variables
  - Front door criterion (never mind about this unless you want a great dissertation project)
- But remember IV! And related methods to analyze quasi-experimental data.

Effect modification in linear models with a continuous variable

In linear models, an interaction between a binary and a continuous variable implies the slope of the line relating the continuous variable to the outcome is different depending on the value of the binary variable

"different slopes for different folks"

$$E(MMSE)=b_0+b_1*HighEd+b_2*Age+b_3*HighEd*Age$$



Interpreting binary/continuous interactions in logistic models

- If you estimate: Logit(p)= ln(p/(1-p)) =–3+.5*X+1.2*Z+.2*X*Z
  - OR_x=exp(.5)=1.65 for Z=0
  - OR_z=exp(1.2)=3.32 for X=0
- X=1, Z=0
  - logit=-3+.5
  - odds=exp(-2.5)=.082
  - p=exp(-2.5)/(1+exp(-2.5))=.076
- X=1, Z=1
  - logit=-3+.5+1.2+0.2=-1.1
  - odds=exp(-1.1)=0.33
  - Odds for X=1, Z=1 / odds for X=1, Z=0= .33/.082=4.02
  - Could also calculate: Exp(1.2+0.2)=exp(1.4)=4.06
- X=1, Z=2
  - logit=-3+.5+2*1.2+2*0.2=0.3
  - odds=exp(0.3)=1.35
  - Odds for X=1, Z=2 / odds for X=1, Z=0= 1.35/.082=16.46
  - Could also calculate: Exp(2*1.2+2*0.2)=exp(2.8)=16.4

Estimation of mixed models: Wide vs Tall Data sets- same issues for people or places or items

- Wide

place	X_pers1	X_pers2	X_pers3
1	27	32	45
2	18	17	20
3	41	44	46

Stata: reshape command
Or w/ arrays

$Y_{ij}=b_{0j}+b1*X_{ij}+b2*W_j+e_{ij}$

Or

$Y_{ij}=b_{0j}+b1_j*X_{ij}+b2*W_j+e_{ij}$

Full specification includes info on the distribution of the random coefficients, e.g., $b_{0j}=b0+m_j$ $\mu_j \sim N(0,\sigma)$

- Tall

Place (j)	person(i)	X_{ij}	W_j
1	1	27	.3
1	2	32	.3
1	3	45	.3
2	1	18	.1
2	2	17	.1
2	3	20	.1
3	1	41	.9
3	2	44	.9
3	3	46	.9

Wide vs Tall Data sets

- Wide

id	X_time1	X_time2	X_time3
1	27	32	45
2	18	17	20
3	41	44	46

Stata: reshape command
Or w/ arrays

- Tall

id	time	X
1	1	27
1	2	32
1	3	45
2	1	18
2	2	17
2	3	20
3	1	41
3	2	44
3	3	46

Wide vs Tall Data sets: Same for people or places or items

- Wide

person	item1	item2	item3
1	27	32	45
2	18	17	20
3	41	44	46

Item 1= Difficulty getting out of bed
Item 2 = Difficulty feeding self
Item 3= Difficulty dressing self

- Tall

Person	Item	X
1	1	27
1	2	32
1	3	45
2	1	18
2	2	17
2	3	20
3	1	41
3	2	44
3	3	46

Classic growth curve: Willis et al Pmed S2: trajectories of SBP

$$SBP_{ij} = \beta_0 + \mu_{0i} + (\beta_1 + \mu_{1i})\, age_{ij} + (\beta_2 + \mu_{2i})\, age^2_{ij} + (\beta_3 + \mu_{3i})\, age^3_{ij} + \varepsilon_{ij}$$

- Where SBP_{ij} is the systolic blood pressure for individual *i* at observation time *j*.
- β_0 and β_1 are the fixed intercept and slope, μ_{0i} to μ_{3i} are the random intercept and slope coefficients for each individual *i*, and ε_{ij} is the residual error term for individual *i*, at time *j*.
- Covariance between the random coefficients was allowed . Age was centred at the baseline age in each cohort .

Classic growth curve: Willis et al Pmed S2: trajectories of SBP

$$SBP_{ij} = \beta_0 + \mu_{0i} + (\beta_1 + \mu_{1i})\, age_{ij} + (\beta_2 + \mu_{2i})\, age^2_{ij} + (\beta_3 + \mu_{3i})\, age^3_{ij} + \gamma_1\, zBMI_{ij} + \gamma_2\, (age_{ij}\, zBMI_{ij}) + \gamma_3\, (age^2_{ij}\, zBMI_{ij}) + \delta_1\, zbHT_i + \delta_2\, (zbHT_i\, age_{ij}) + \delta_3\, (zbHT_i\, age^2_{ij}) + \delta_4\, (zbHT_i\, age^3_{ij}) + \varepsilon_{ij}$$

- $zBMI_{ij}$ is the BMI z-score externally referenced to the UK 1990 sex and age-specific growth charts for subject *i* at observation time *j*.
- γ_1 represents the cross sectional association between BMI and SBP
- γ_2 and γ_3 allow the association between current BMI and SBP to vary by chronological age.
- $zbHT_i$ represents the height z-score at baseline referenced to the UK 1990 sex and age-specific growth charts for subject *i*.
- δ_1, δ_2 and δ_3 thus allow baseline height to affect the SBP intercept and slope.
- Non significant γ's and δ's were removed from the final model.

What's the best time dimension?

- Age is conceptually natural (?)
  - But using age as the time-dimension conflates between person differences and within person changes
  - For studies of old people, cohort effects can bias age-based growth curves
  - Especially problematic if exposure of interest is highly correlated with age
- Using "time from baseline" and adjusting for baseline age eliminates cohort problems, but introduces potential period and practice effect problems
- Best practice: estimate separately (baseline age + time from baseline) and understand why they diverge (Fitzmaurice, Laird, and Ware, pg 420)

Alternative time dimensions

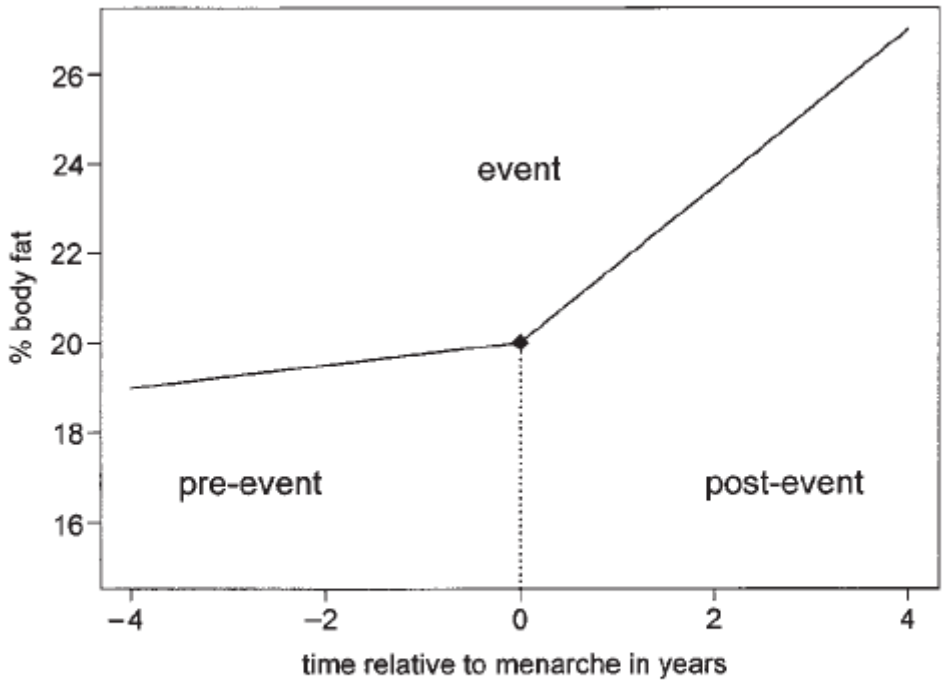


Figure 4 Graphical representation of the research hypothesis and modelling approach: the piecewise model assigns a pre-event slope β_1 and a post-event slope β_2 separated by the menarcheal event

Choose a time-dimension:

- Age
- Time before or after a major event
- Secular time

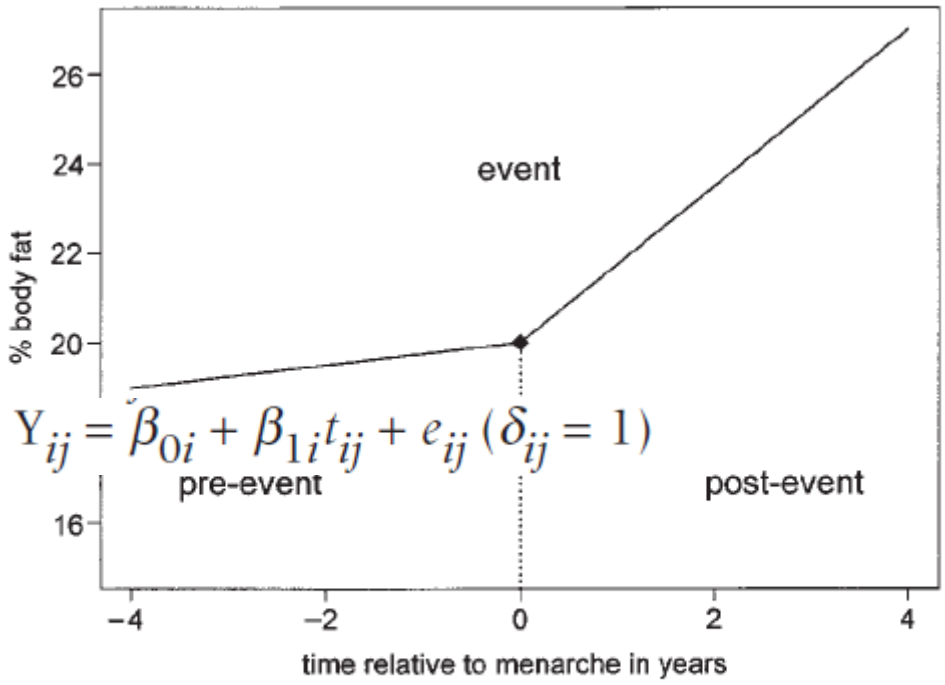
Specify a model with time as a linear predictor

- How does body fat change per year of age?

Add a factor that modifies the rate of change with time

- Is the rate of change different before (δ=1) or after (δ=0) menarche?

First the model for before menarche (δ=1)



$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij} \; (\delta_{ij} = 1)$$

Figure 4 Graphical representation of the research hypothesis and modelling approach: the piecewise model assigns a pre-event slope β_1 and a post-event slope β_2 separated by the menarcheal event

Choose a time-dimension:

- Age
- Time before or after a major event
- Secular time

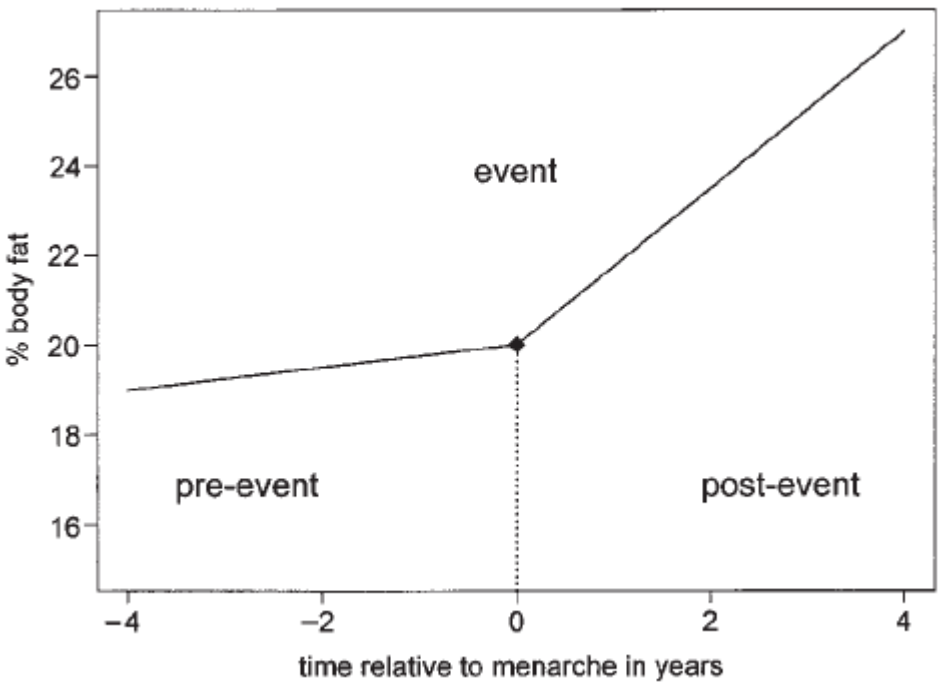
Specify a model with time as a linear predictor

- How does body fat change per year of age?

Add a factor that modifies the rate of change with time

- Is the rate of change different before (δ=1) or after (δ=0) menarche?

Provides a flexible test of multiple questions



Does rate of change differ before or after menarche?

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} \delta_{ij} + \beta_{2i} t_{ij} (1 - \delta_{ij}) + e_{ii},$$

Does rate of change differ after menarche depend on parental obesity (v_i)?

$$Y_{ij}=\beta_0+\beta_1 t_{ij}\delta_{ij}+\beta_2 t_{ij}(1-\delta_{ij})+\beta_3 v_i+ \beta_4 t_{ij}(1-\delta_{ij})v_i+\varepsilon_{ij}$$

This equation estimates a slope for pre-menarche and post-menarche, but lets the post-menarche slope differ by parental obesity

Sources of uncertainty: ways to address

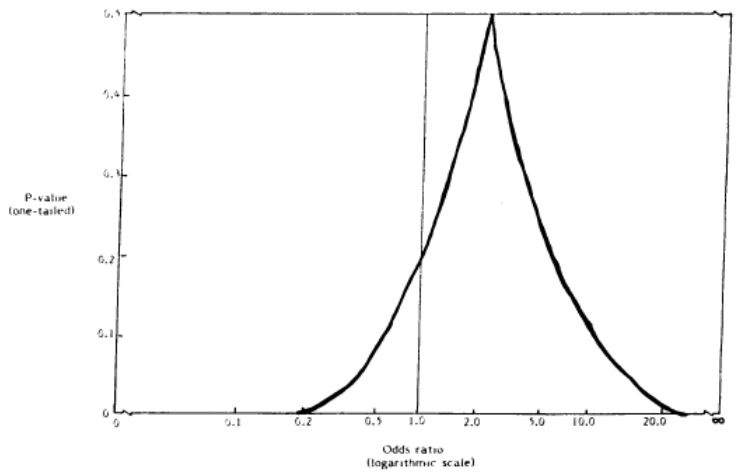
- Is your code doing what you think it is doing? Code review
- If you repeated this again with all exact same sampling procedure, but draw a new sample, and apply identical analysis procedures, what would be the distribution of this parameter? Confidence intervals, simulations
- If I repeated this again, with identical data but left it to a reasonable scientist to make analytic decisions s/he thought best? Post your code, independent replication
- If I repeated this again in a new setting with different populations, measurement instruments, scientists… would I find the same result? Transportability+independent replication, meta-analyses
- If I tried to test the theory in a whole new way, would I find the same result? Triangulation. IV, different confounding structure, negative controls, testable implications

Jean's questions

- Can you explain slide 44 of uncertainty lecture on relationship between CIs and p-values?

P-values: some common traps to avoid

- The p-value > .05, therefore, $\beta=0$ (don't infer this)



We usually test against the null hypothesis, eg $\beta=0$
But conceptually, you could test the probability you would get the estimate you actually did under any alternative, e.g., are my data consistent with the possibility that $\beta=-1$?

As you test against null values that are closer and closer to what you actually estimated in your sample, i.e., β=estimated(β), the p-value should approach 1.
What you actually estimated is the coefficient *most consistent* with your data.

From: Poole, Beyond CIs, 1987

Confidence Intervals

- Suppose your estimated(β)=1.5
- Imagine you tested the null that β=estimated(β)–2=–.5 & p=.001
  - i.e., you'd only estimate β=1.5 if 1 time out of a 1000 times if actual β=-.5
- and then you test the null that β=estimated(β)–1=.5 and p=.009
  - i.e., you'd only estimate β=1.5 if 9 times out of a 1000 times if actual β=.5
- and then you test the null that β=estimated(β)–.5 and p=.06
  - i.e., you'd only estimate β=1.5 if 6 times out of a 100 if actual β=1.0
- You can define the CI this way:
  - The set of parameter values for which the P-value exceeds some threshold, e.g., 0.05

Confidence Intervals (slide 44)

- You can define the CI this way:
  - The set of parameter values for which the P-value exceeds some threshold, e.g., 0.05
  - Calculate the P-value comparing the observed data to a particular "truth" (e.g., beta=0, 0.1, 0.2, 0.3…4.0, 4.1)
  - Identify the range of possible betas for which the P-value (for testing your actual estimate against this possible beta) is >0.05
  - That's the 95% confidence interval for the beta
  - A 95% CI will over unlimited repetitions contain the true parameter at least 95% of the time (if statistical model is correct and there's no bias)

Confidence Intervals

- A 95% CI will over unlimited repetitions contain the true parameter at least 95% of the time (if statistical model is correct and there's no bias)
- CI gives you more information on how much you've learned from the study compared to just a p-value
  - Very wide CI: not much info.
  - i.e., large SE and large variance
  - Sometimes show SE instead of CI
- Often the CI is just used as significance tests: if the CI doesn't include the null, P<.05 for test of the null. This is ignoring advantages of the CI.

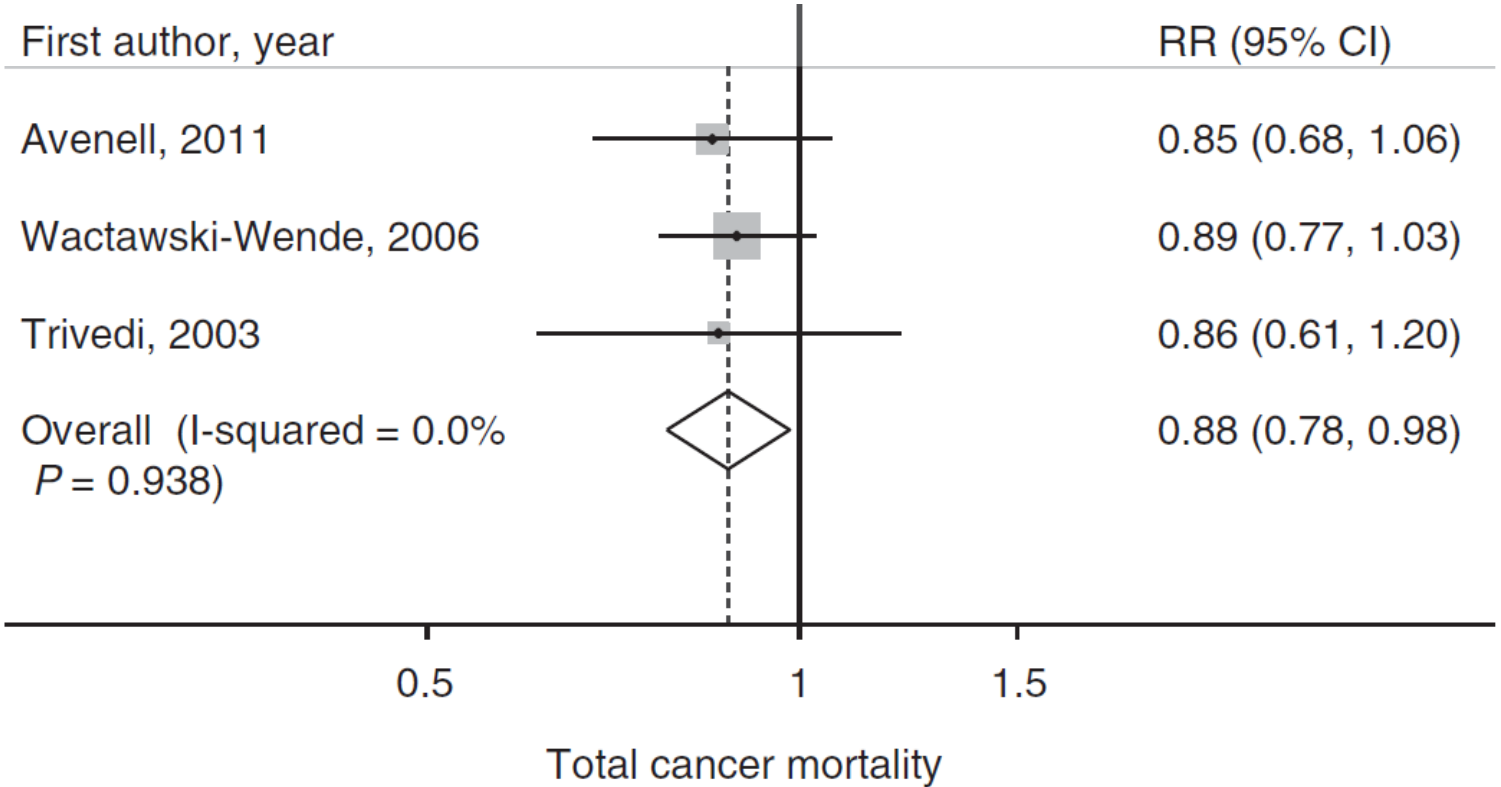
Meta-analyses

- Meta-analyses can help with three categories of uncertainty
  - Sampling variability because they effectively increase the sample size
  - Researcher judgment uncertainty to extent that researchers are independent
  - New setting uncertainty because often conducted with slight variations in population, measures, etc.

Meta-analysis of RCTs of vitamin D supplements and cancer mortality

Meta-analysis goal is to combine effect estimates usually weighting by the inverse of the variance (ie the information in the study).

No individual study convincing, but 3 studies, with nearly identical RRs, fairly convincing evidence



Meta-analysis: fixed effects model vs random effects?

- Use fixed effects if you believe the only reason effect estimates differ between studies is chance due to sampling variability for those enrolled in that study:
  - All studies are functionally identical (because study design would lead to differences in effect estimates
  - There is a single "true" effect size that prevails in all populations to which you are generalizing.
- Fixed effects model assumes the observed effect (β_i) for any study is:
  - $\beta_i=\theta+\varepsilon_i$
- Where θ is the true effect size, which is the truth for all study settings
- And ε_i is the difference between θ and the effect in this study, which is estimated by the standard error for the effect estimate in study *i*
- Each study is weighted (W_i) by the inverse of its variance: $V_i=\frac{\sigma^2}{n}$
  - Big variance→ imprecise study → count it less
- Calculate the pooled effect as $M=\frac{\sum_i^k W_i\beta_i}{\sum_i^k W_i}$

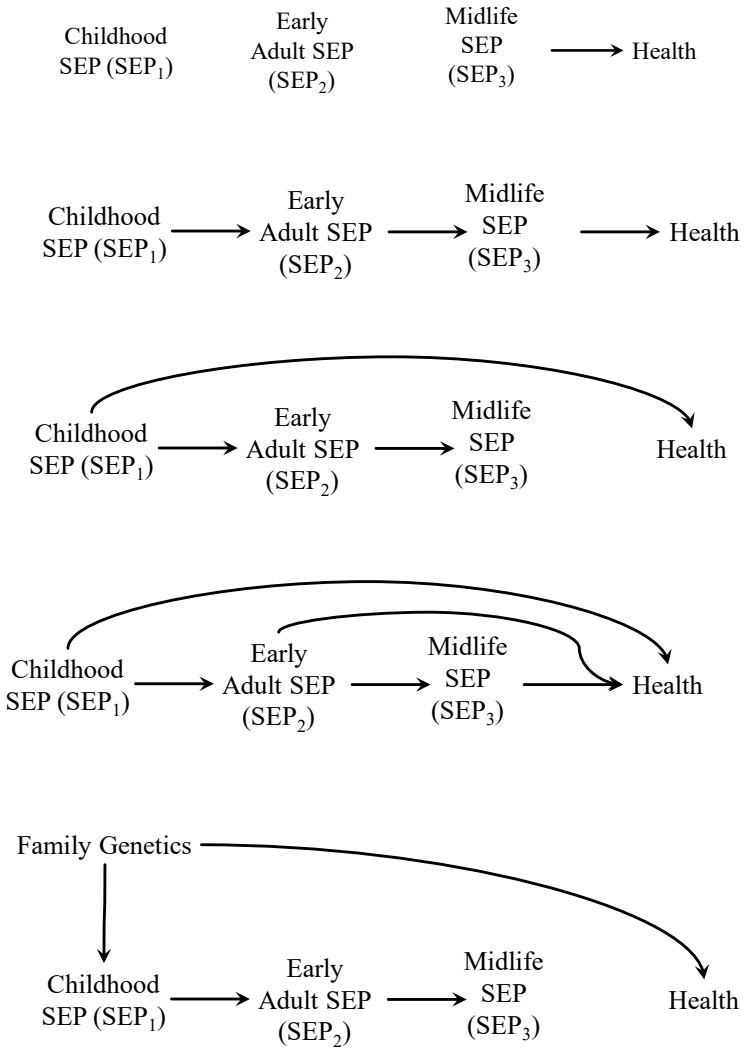
Use random effects if effect estimates differ between studies due to both sampling variability within studies and differences in the true effect in each study.

- Random effects model assumes the observed effect (β_i) for any study is: $\beta_i=\mu+\zeta_i+\varepsilon_i$
- Where μ is the grand mean of effects across all studies
- ζ_i is the difference between the grand mean effect and the effect in study *i*
- ε_i is the sampling error for the effect estimate in study *i*, i.e., the difference between the true mean for study *i* and the observed mean for study *i*.
- Each study is weighted (W_i) by the inverse of *both* sources of variance: $W_i=\frac{1}{V_i+T^2}$
  - Study specific variance + Between study variance
- Calculate the pooled effect the same as for fixed effects: $M=\frac{\sum_i^k W_i\beta_i}{\sum_i^k W_i}$
- Trick is estimating the between-study variance (intuition: how much between study variance would be expect due to sampling? The rest is due to true effect heterogeneity)

Random, random everywhere…

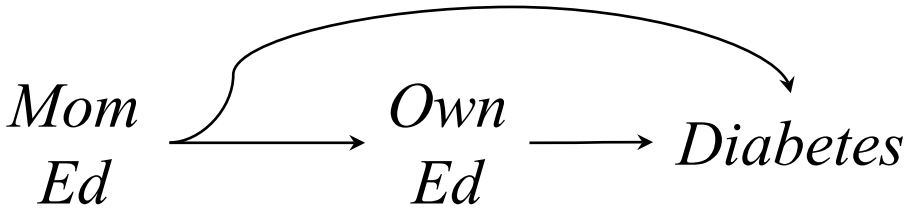
- Remember our other uses of random effects models, where we used random deviations from an overall average to indexed clusters like a specific person (with repeated measures on that person) or a specific neighborhood (with multiple people observed within that neighborhood)?
  - Clusters with fewer observations got lower weight
  - We assumed that the overall average (random intercepts) or average effects (random slopes) differed between clusters
  - Best predictions for "truth" in any given cluster was a weighted average of the overall average and that particular cluster's values.
  - We asked questions about both the overall average effect estimate and about the variance between clusters.
- One other thing to remember for later (multiple imputation/Rubin's rules).
  - Each study is weighted (W_i) by the inverse of *both* sources of variance: $W_i=\frac{1}{V_i+T^2}$
  - Because the total uncertainty is the sum of: Study specific variance + Between study variance

Mediation questions turn up everywhere. Lifecourse epi is one example but common if you cannot directly intervene on an exposure.



- To distinguish *latency, cumulative, or trajectory* from *immediate* risk:
- E(health)=b0+b1*SEP1
- If you have some measured confounders (common causes or factors on a common cause pathway):
- E(health2)=b0+b1*SEP1+b2*Age +b3*Birth Regn
- Under immediate risk, b1=0

Mediation: Classic Decomposition Approach

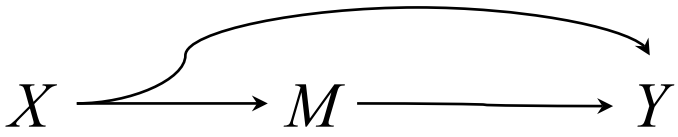


- Total effect= direct effect + indirect effect
- Indirect effect= total effect-direct effect
- Total effect= E(Db | MomEd=1)-E(Db|MomEd=0)
- Direct effect=E(Db|MomEd=1, OwnEd)-E(Db|MomEd=0, OwnEd)
- Sometimes called “Barron-Kenny” decomposition

Assumptions for classic approach

1. No confounding of exposure-outcome
2. No confounding of mediator-outcome
3. No confounding of exposure-mediator
4. No interaction between mediator and the direct pathway
5. No measurement error of the mediator
6. Linear models

Mediation: Counterfactual Definitions of Direct vs Indirect Effects Decomposition



- Controlled direct effect: The difference in counterfactual Y if you set X to 1 vs X to 0 but in both circumstances also set M=0
  - $Y_{x=1,m=0}$ and $Y_{x=0,m=0}$
- Also a controlled direct effect: The difference in counterfactual Y if you set X to 1 vs X to 0 but in both circumstances also set M=1
  - $Y_{x=1,m=1}$ and $Y_{x=0,m=1}$
- These two don't have to be the same, so there are as many values of the "controlled direct effect" as there are possible values of M.
- Controlled "indirect" effect doesn't make sense because we are setting M to the same value for everyone in the controlled framework.

Natural Direct and Indirect Effects: Logistic Model

- Natural direct effect: The difference in counterfactual Y if you set X to 1 vs X to 0 but in both circumstances also set M to what it would be for a reference value of X, e.g., X=0
- M will differ between different people, but theoretically unrelated to X
- Sometimes criticized because it requires a cross-world contrast: the potential outcome Y if X=1 but M= what it would be if X=0
- If no X-Y, X-M, or M-Y confounding, no interaction between X and M, and rare outcome, then classic BK is approximately NDE/NIDE
  - $\text{Logit}(p(Y|X,M,C))=a_0+a_1*X+a_2*M+a_3*C$
  - $E(M|x,c)=b_0+b_1*X+b_2*C$
  - $\text{NDE}=\text{CDE}=\exp(a_1*(x-x'))$
  - $\text{NIDE}=\exp(b_1*a_2*(x-x'))$

Natural Direct and Indirect Effects: Logistic Model

- If interaction between X and M estimate:
- $\text{Logit}(p(Y|X,M,C))=a_0+a_1*X+a_2*M+a_3*X*M+a_4*C$
- $E(M|x,c)=b_0+b_1*X+b_2*C$

- $CDE(m)=\exp((a_1+a_3*m)*(x-x'))$
- $NIDE=\exp((a_2*b_1 + a_3*b_1*x)(x-x'))$

Controlled vs Natural Direct Effects

- CDE requires fewer assumptions
- NDE cannot be estimated with recanting witness that is influenced by X and a common cause of M and Y.
- Debate about which is more policy relevant
- Often not that different, so debate is perhaps overblown, but check for interactions

Jean's questions

- Why for natural direct and indirect effect do we always set M to what if would be for the reference value of X (X = 0) rather than some other value of X (e.g. X = 1)?
  - You can set it to either. There are special names for these two, but we say in NDE we set the mediator to what it would have been "in the absence of the exposure" implying it is when X=0.
- Can you review why we would want the controlled vs. natural effects in mediation?
  - What is a plausible intervention? Could you set everyone to have a particular value of the mediator? Or could you simply interrupt the effect of X on M?
  - Do you want to decompose the total into the direct and indirect? Use natural.
  - Do you meet the assumptions for estimating the natural?
- Is this correct: If there is no effect modification between the indirect pathway and the direct pathway, then there is only one CDE and it is equal to the NDE?
  - Yes

Estimation

- Now lots of packages, but if you have the formulas, you can always calculate what you need, so don't get too tied to a particular software package.
- Imai's generalized approach:
  - To estimate any direct or indirect effect, just estimate the potential outcomes.
  - Let $Y_{x=1,m=1}$ represent the counterfactual value of Y setting X to 1 and M to 1
  - Let $Y_{x=1,m(x=1)}$ represent the counterfactual value of Y setting X to 1 and M to whatever value M would take if X were set to 1.

Estimation

1. Fit a model for the mediator, as a function of exposure and confounders
   - For example reg M X C
   - Make two fake-world copies of each observation, but set X to 0 in one world and X to 1 in the other
   - Generate fake-world predictions using the regression equation from the real world
   - So you have predicted values of $M_{x=0}$ and $M_{x=1}$
2. Fit a model for the outcome, as a function of exposure, mediator, and confounders (if you need to control for variables you cannot include in the outcome model, use IPW). Remember to allow for exposure*mediator interactions if those are plausible.
   - For example reg Y M X M*X C
   - In your fake world where you used X=0 to predict M, use the outcome regression equation to predict the value of Y when X=0 and $M=M_{x=0}$
   - In your fake world where you used X=1 to predict M, switch everyone to X=0 and use the outcome regression equation to predict the value of Y when X=0 and $M=M_{x=1}$
   - So you have predicted values of $Y_{x=0}$ and $M_{x=1}$
3. Now you have $Y_{x=0,m=m(x=1)}$ and $Y_{x=0,m=m(x=0)}$ and the difference between these two is the indirect effect of X on Y mediated by M when X is 0, i.e., the Natural indirect effect.
4. Bootstrap the Cis.

Estimation

- Now you have $Y_{x=0,m=m(x=1)}$ and $Y_{x=0,m=m(x=0)}$ and the difference between these two is the indirect effect of X on Y mediated by M when X is 0, i.e., the Natural indirect effect.
- You can do this for any definition of treatment vs no treatment,
- Simulate the potential values of the mediator under treatment versus no treatment (whatever contrast you would like, e.g., "all treated" vs "nobody treated" or "everyone has education=12" vs "everyone has education=8 years" or "one third of the population has educ=16, one third educ=12, and one third=8" vs "one half has educ=12 and one half has educ=16").
- You can also do this to calculate $Y_{x=1,m=0}$ and $Y_{x=0,m=0}$ the controlled direct effect setting M to 0 or $Y_{x=1,m=1}$ and $Y_{x=0,m=1}$ the controlled direct effect setting M to 1

Other approaches

- Inverse Odds Ratio Weighting
  - Use the mediator to predict the exposure first, and use the odds ratio to characterize that association because the odds ratio is the same even if you switch exposure-outcome.
  - So M->X is the same as X->M
  - Now calculate the inverse of the odds that the mediator took the value it did, given the value of X for each individual.
  - In the weighted population, X has no effect on M, so calculate the effect of X on Y in this weighted population to estimate the direct effect of X on M
  - Uses the same IPW trick!
  - Very clever. Convenient for multiple mediators. Convenient for any functional form of the outcome model.
  - Weights lead to imprecise estimates.

Jean's questions

- Can you explain slide 44 of uncertainty lecture on relationship between CIs and p-values?

- Why for natural direct and indirect effect do we always set M to what if would be for the reference value of X (X = 0) rather than some other value of X (e.g. X = 1)?

- Can you review why we would want the controlled vs. natural effects in mediation?

- Is this correct: If there is no effect modification between the indirect pathway and the direct pathway, then there is only one CDE and it is equal to the NDE?

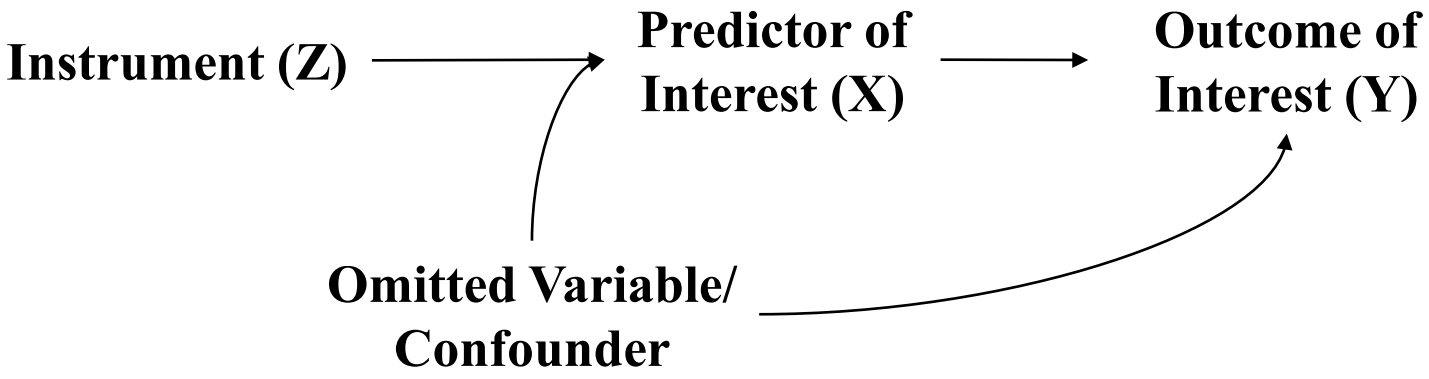
What is the difference between sufficient-cause interaction and epistatic interaction?

- A "sufficient cause interaction" is present if there are individuals for whom the outcome would occur if both exposures were present but would not occur if just one or the other exposure were present.
  - Sufficient cause does not require $Y_{00}=0$
- There is a "singular" or "epistatic" interaction if there are individuals in the population who will have the outcome if and only if both exposures are present; in counterfactual notation, that is, there are individuals for whom
  - $Y_{11} = 1$ but $Y_{10} = Y_{01} = Y_{00} = 0$.

Instrument Assumptions: in intuitive terms

Z is an instrument for the influence of X on Y if:

1. Z predicts X
2. Z has no effect on Y unless the effect is mediated by X
3. No other variables influence both Z and Y
4. X does not affect Z (redundant but people are often confused about this)



Obtaining IV Estimates

Wald Estimates:

$$\text{Effect} = \frac{\text{(Difference in Avg Outcome between IV levels)}}{\text{(Difference in Avg Exposure between IV levels)}}$$

Two Stage Least Squares Estimates:

$E(X)=\beta_0+\beta_1 IV + \beta_k$ Other Covariates

Outcome$=\gamma_0+\gamma_1 E(X) + \gamma_k$ Other Covariates

γ_1 is a consistent estimate of the causal effect (e.g., LATE, ATE, ETT, depending on assumptions)

Variations on IV Estimators

Wald Estimate: $= \frac{\text{(Difference in Avg Outcome between IV values)}}{\text{(Difference in Avg Treatment between IV values)}}$

Two Stage Least Squares (2SLS):

- Multiple instruments
- Control for covariates

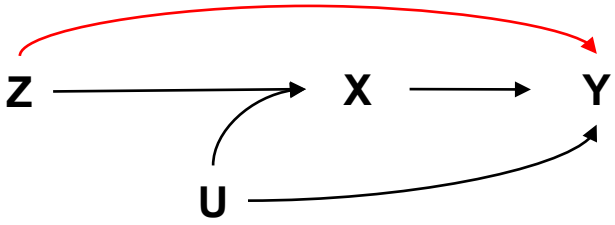
Separate Sample IV:

- 1st and 2nd stage of a 2SLS are from different data sets
- Often relevant in MR: assumes 2 samples are "equivalent"

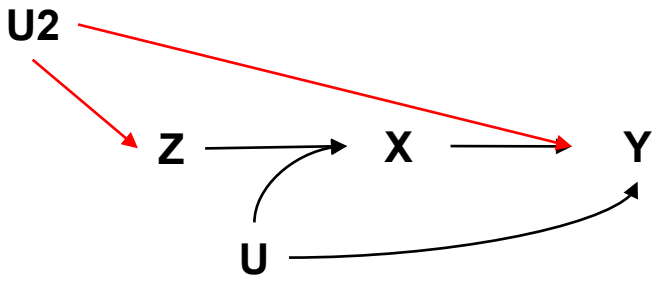
- Residual control approaches

Doubting Instruments: major biases similar to critiques of RCTs

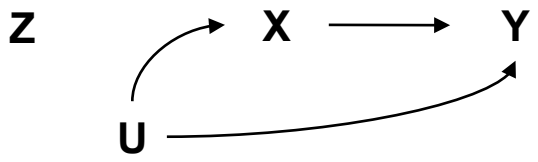
- Do they have other pathways to the outcome?
  - *Unblinded trials: controls become demoralized*



- Is there a common cause of the instrument and the outcome?
  - *Trials: unfair random assignment*



- Do they actually affect anyone's exposure?
  - *Trials: nobody adheres to assignment*



Improving IV Studies

- Stronger IVs with larger effects on treatment
  - In MR: better genetic risk scores for phenotypes
  - In PP: larger practice effect differences, stronger encouragements
- Control for most plausible violations
- Sensitive tests for assumptions
- Sample size improves situation because tests of violations are informative and allows for more control variables.

Evaluating the assumptions

1. Constraints implied by theory
2. Over-identification tests
3. IV inequality constraints
4. Stratification-based tests (similar to over-identification)
5. Negative controls
6. Independent from known confounders

1 Evaluating MR Studies: leverage prior assumptions about confounding

- Compare IV effect estimate to conventional effect estimate: if conventional effect estimate is positively confounded, IV<conventional
- Relies on assumption regarding the direction of confounding: if you don't know the direction, the test is not informative
- 4 *equivalent* versions of this test: no more convincing to show all 4.
- Often, field is only interested in knowing if a conventional effect estimate is biased up
- Other direction of bias not of interest, or not considered plausible
- Not guaranteed to be consistent in non-linear causal structures (but doesn't completely rely on linearity).

1 Evaluating MR Studies: leverage prior assumptions about confounding

1. If the conventional estimate is biased upwards, and the IV is unbiased, the IV estimate should be smaller than the conventional estimate
2. If the IV effect is fully mediated by the phenotype, adjusting the regression of the outcome on the instrument for the phenotype should attenuate the effect estimate for the instrument
3. Adjusting a biased conventional estimate for a valid instrument *exacerbates* the bias
4. The residual between the predicted value of the outcome based on the IV estimate and the actual value of the outcome is correlated with the phenotype.

2 Evaluating IV Studies: Over-identification tests

- Use multiple instrumental variables to conduct over-identification tests.
  - Physician preference versus regional preference versus formulary changes
  - In MR: Other genes.
- Cannot detect violations of the IV assumptions if all instruments have identical biasing pathways.
- May also reject even when all instruments are valid if the model is incorrectly assumed to be linear or the phenotype is composite. These tests generally have low statistical power.

Limitations of Over-ID Tests

- Power challenges: lots of uncertainty around 2 IV estimates will fail to reject
- May reject even if both instruments are valid, but have different subpopulations of "compliers", with different causal effects of X on Y (i.e. different LATEs)
- It's hard to get just 1 good instrument, much less 2 or 3.
- Data from GWAS may help - debatable

3 Evaluating IV Studies: instrumental inequality tests

- These tests are applicable only when the causal variable is known to be categorical (drug, no drug).
- Certain inequalities are impossible given the IV assumptions: if you see them, IV must not be right

$$\max_i [\Pr(X=0, Y=1|G=i)] \leq \min_i [1 - \Pr(Y=0, X=0|G=i)]$$

$$\max_i [\Pr(X=0, Y=1|Z=i) + \Pr(X=1, Y=1|Z=i)] + \max_i [\Pr(X=0, Y=1|Z=i) + \Pr(X=1, Y=0|Z=i)] + \max_i [\Pr(X=0, Y=0|Z=i)] \leq 2$$

$$\max_i [\Pr(X=1, Y=1|Z=i)] \leq \min_i [1 - \Pr(Y=0, X=1|Z=i)]$$

- Detects extreme violations of the assumptions
- With more instrumental variables, more opportunities for violations of tests.

4 Evaluating IV Studies: Modifying factors

- Identify factors that modify the IV-treatment association.
- Compare the IV effect estimate across groups in which the population association between the instrument and the treatment is either silenced or reversed.
- This test could identify a biased instrument if the biasing pathway is active in both subgroups.

Stratification Tests: Gene Environment Interactions

- An environmental context changes the relationship between the genotype and the phenotype (the first stage of the IV analysis).
  - "Silencing": genotype and phenotype are independent in some environments.
  - "Modifying": genotype changes the probability of the phenotype by a different extent in different environmental contexts.
  - "Reversing": genotype increases probability of the phenotype in some environmental contexts and *decreases* probability of the phenotype in other environments (violations of monotonicity).

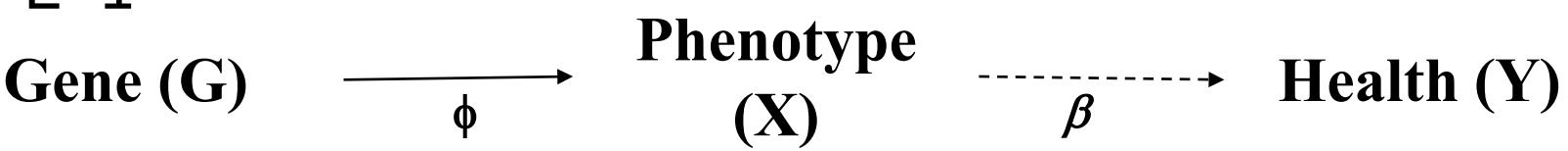
How can this help with MR?

- Assumptions of MR are untestable when they stand alone.
- Combined with additional causal knowledge, these assumptions have testable implications.
- GxE interactions add new causal knowledge – may be able to tell whether associations are inconsistent with hypothesized MR-valid causal structure.

MR Stratified by Gene Silencing Environmental Characteristic

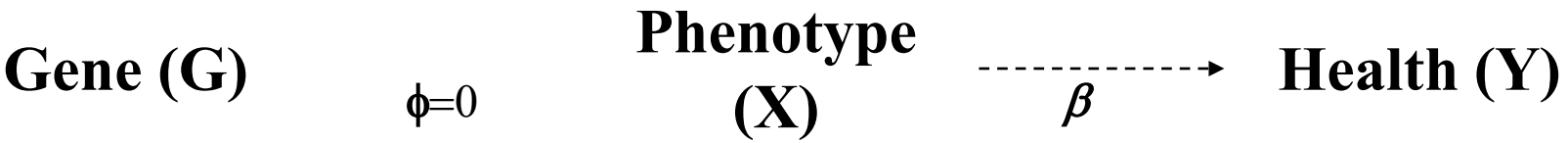
If unbiased:

E=1



$\beta_{MR} = \phi * \beta / \phi = \beta$

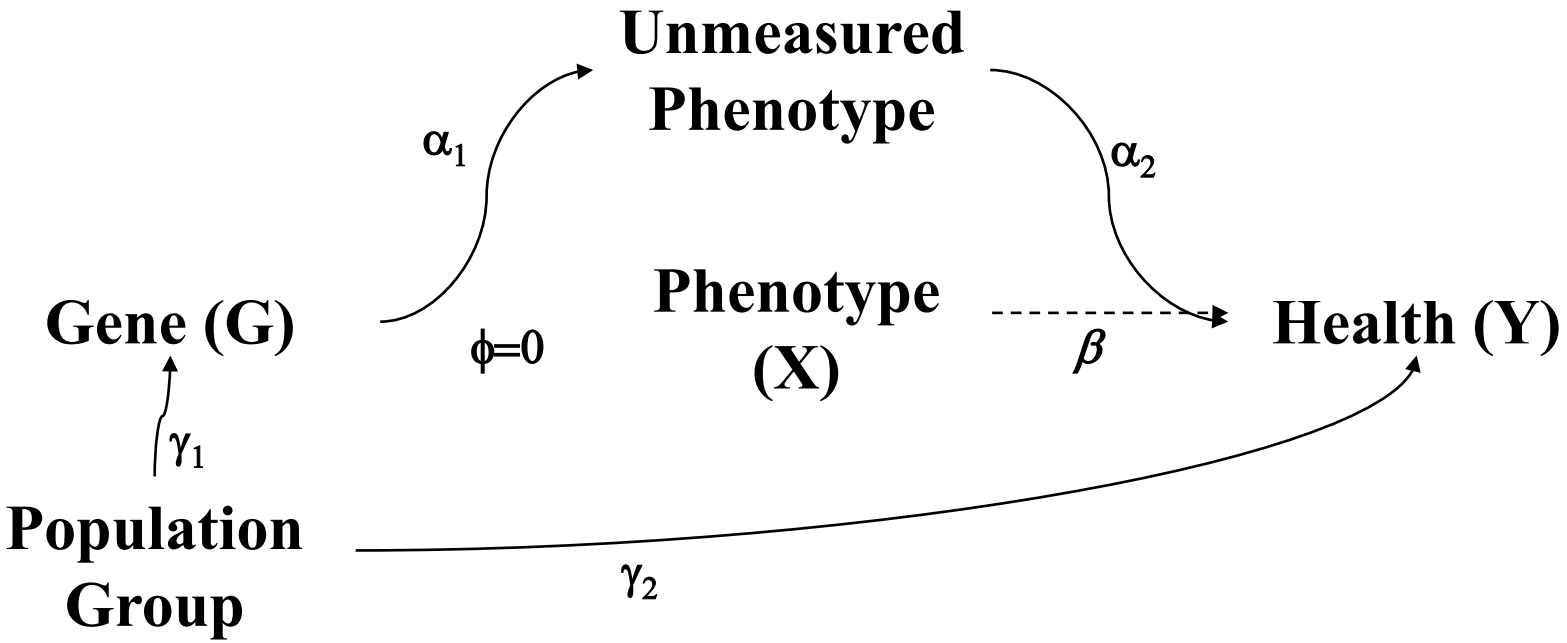
E=0



$\beta_{MR} = \phi * \beta / \phi = 0/0$

MR Stratified by Gene Silencing Environmental Characteristic

E=o (biased case)

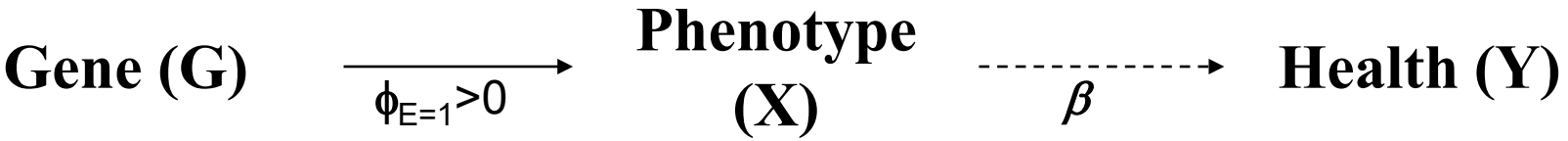


$$\beta_{MR} = (\phi\beta + \gamma_1\gamma_2 + \alpha_1\alpha_2) / \phi = (\gamma_1\gamma_2 + \alpha_1\alpha_2) / 0$$

MR Stratified by Gene Reversing Environmental Characteristic

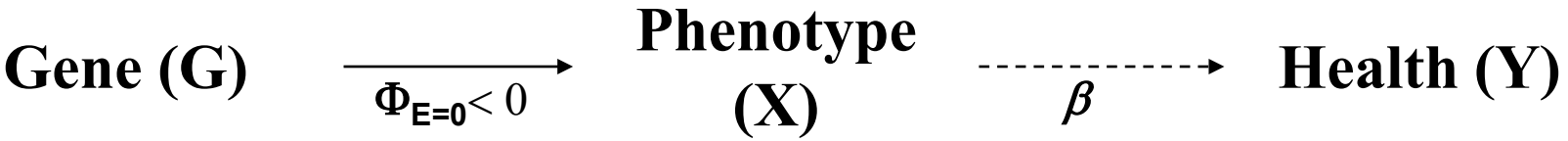
If unbiased:

E=1



$\beta_{MR} = \phi_{E=1} * \beta / \phi_{E=1} = \beta$

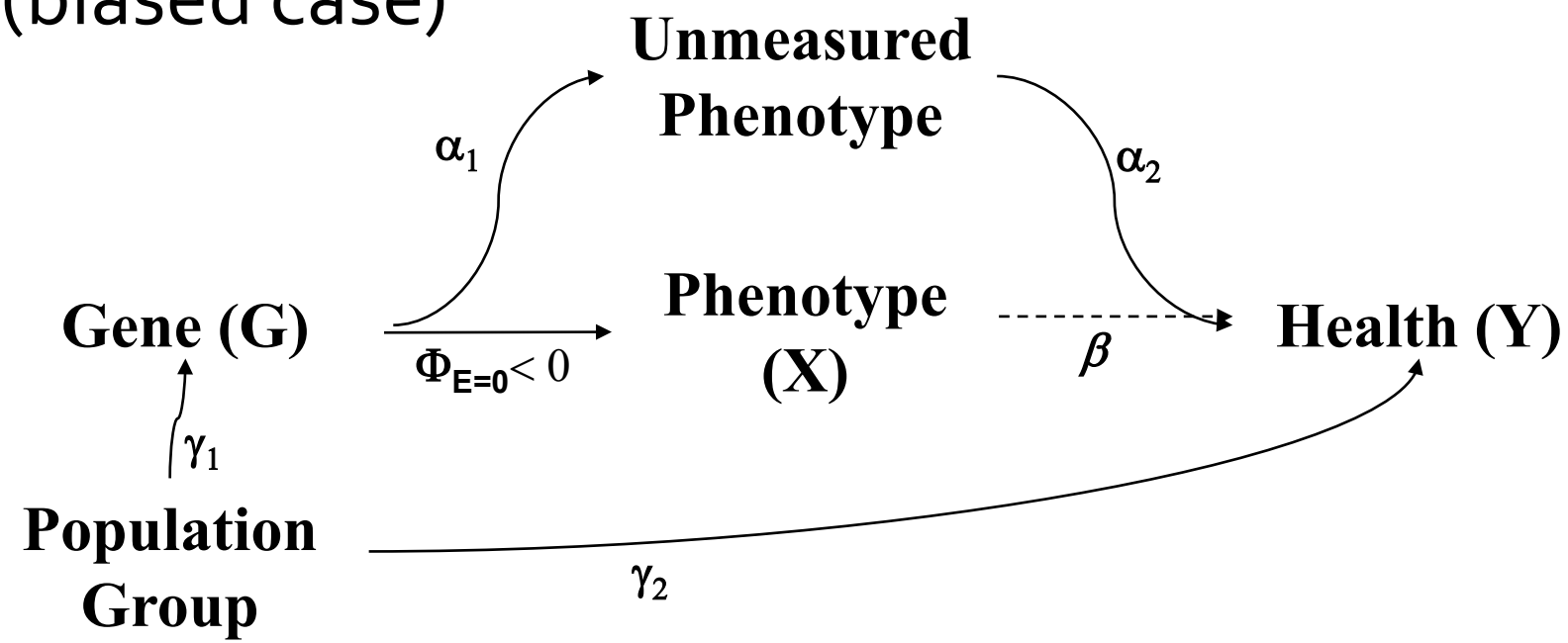
E=0



$\beta_{MR} = \phi_{E=0} * \beta / \phi_{E=0} = \beta$

MR Stratified by Gene Reversing Environmental Characteristic

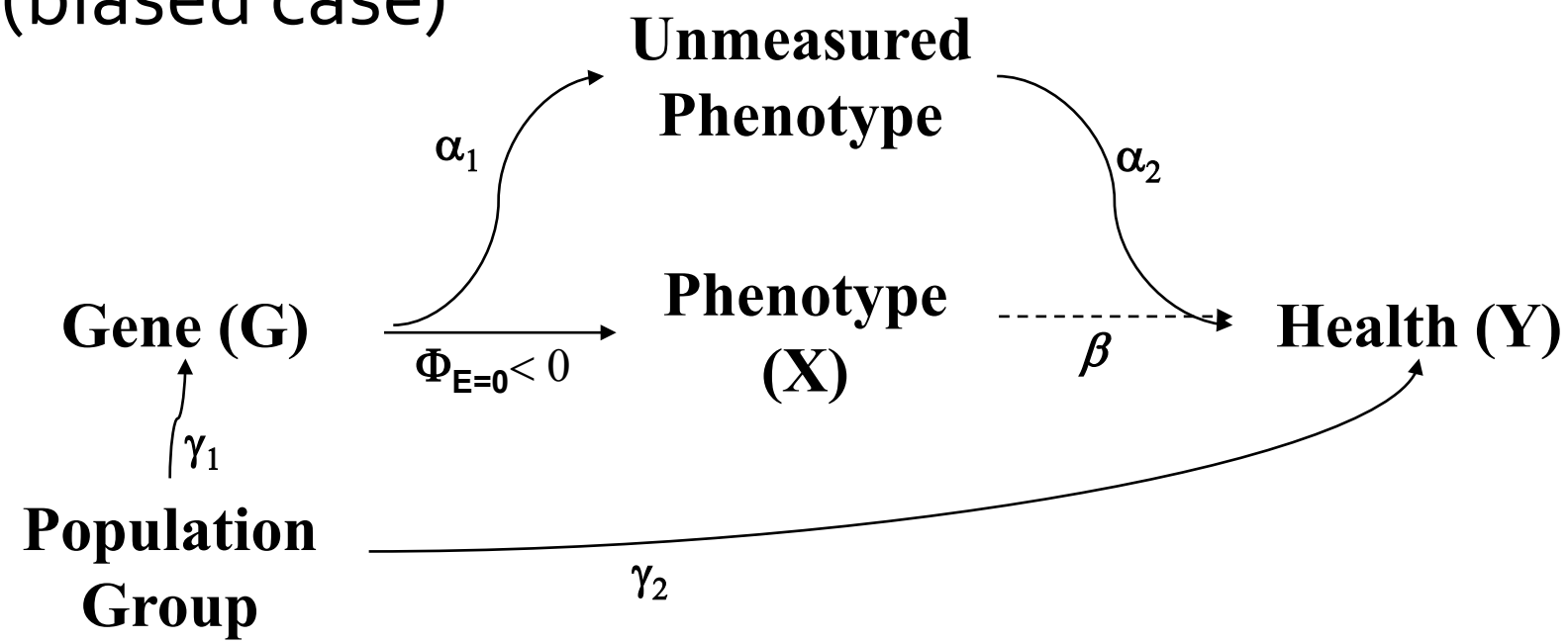
E=o (biased case)



$$\beta_{MR} = (\phi_{E=o}\beta + \gamma_1\gamma_2 + \alpha_1\alpha_2)/\phi_{E=o}$$

MR Stratified by Gene Reversing Environmental Characteristic

E=o (biased case)



$$\beta_{MR} = (\phi_{E=o}\beta + \gamma_1\gamma_2 + \alpha_1\alpha_2)/\phi_{E=o} \neq \beta \neq (\phi_{E=1}\beta + \gamma_1\gamma_2 + \alpha_1\alpha_2)/\phi_{E=1}$$

5 Evaluating IV Studies: Negative controls

- Identify outcomes with similar confounding structure but that are definitely not affected by exposure.

6 Evaluating IV Studies: Independence from measured confounders

- Identify measured confounders of exposure-outcome association
- Test whether these variables are independent of the IV
- If not, you can control for these confounders, but must worry about unmeasured confounders
- Sometimes most coherent to examine independence conditional on a core set of confounders, e.g., age, population eigenvectors