Introduction to Genetic Instrumental Variables or "Mendelian Randomization" Studies

Maria Glymour
Department of Epidemiology & Biostatistics
University of California, San Francisco

Outline

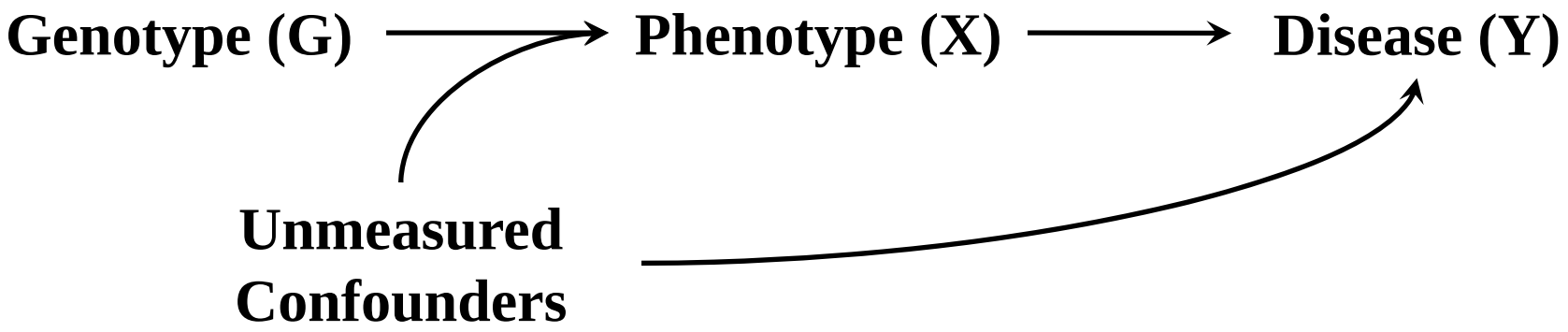
- Mendelian Randomization designs: identifying non-genetic disease determinants
  - Motivation
  - IV and MR assumptions and violations
  - Interpretation of the MR effect estimate
  - Evaluating the assumptions
- Extensions/Future Directions for MR
  - Poly/multi-genic
  - GxE
  - Lifecourse

Some discipline specific (?) assumptions

- Questions of interest are about "causation": "would intervening on exposure X change the value of outcome Y?" i.e. $Y_{X=1}$ vs $Y_{X=0}$
- Assume we never know this at the individual level, because we never find out what Y would have been if exposure had been different than it actually was
- Unfortunately $Y_{X=1} \neq Y|X=1$ because the people with X=1 may also tend to have some other feature, say U, that influences the value of Y.
- U is a confounder: a common prior cause of X and Y.
- We can estimate the effect of X on Y at a group level by randomization, pseudo-randomization, or controlling for common causes of X and Y.

Mendelian Randomization Design:
pseudo-randomization

Use genotype as an *instrument* to estimate the effect of a phenotype on the outcome.



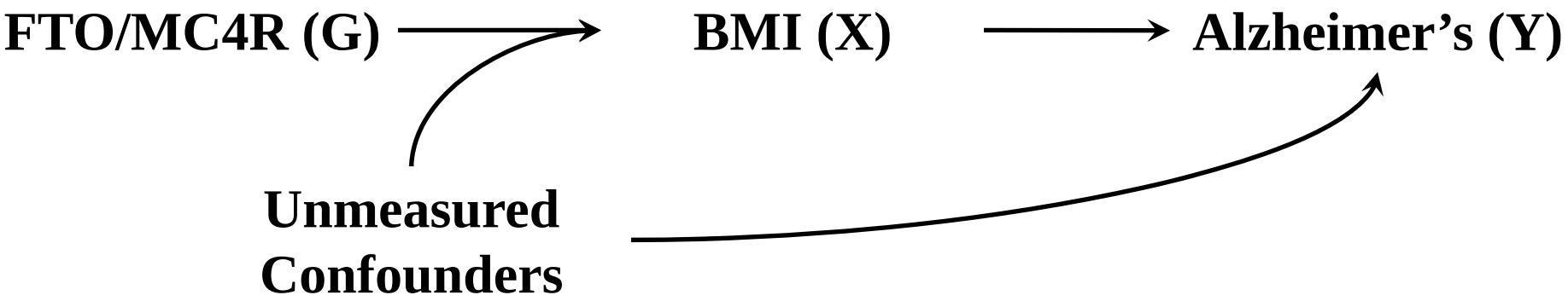
MR is a special case of the long established econometric method of *instrumental variables analysis*.

Appealing when causality is uncertain:

1) Unavoidable confounding
2) Reverse causation
3) Uncertainty: biomarker/endophenotype or cause

Mendelian Randomization Design

Use genotype as an *instrument* to estimate the effect of a phenotype on the outcome.



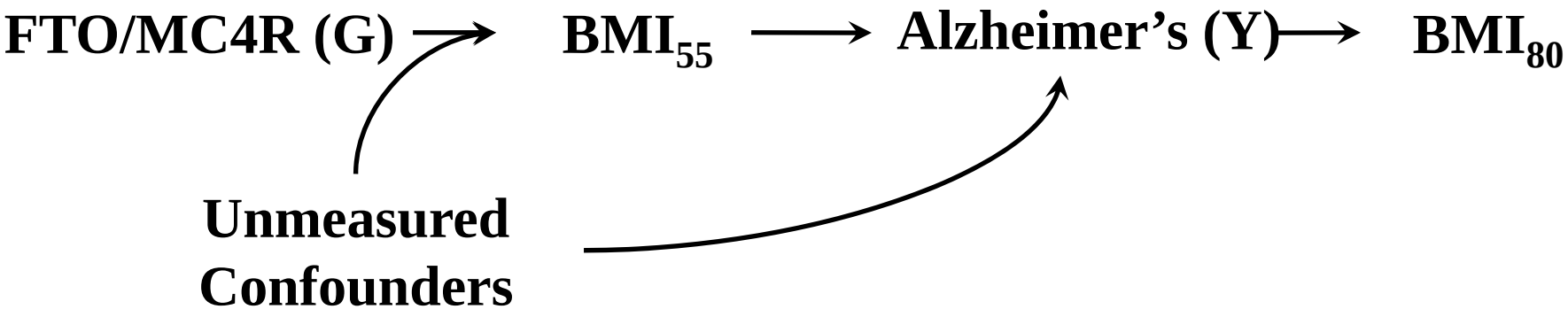
MR is a special case of the long established econometric method of *instrumental variables analysis*.

Appealing when causality is uncertain:

1) Unavoidable confounding
2) Reverse causation
3) Uncertainty: biomarker/endophenotype or cause

Mendelian Randomization Design

Use genotype as an *instrument* to estimate the effect of a phenotype on the outcome.



For age-related outcomes, MR is appealing because we think incipient disease affects many of the "exposures" of interest, i.e. "reverse causation".

Mendelian Randomization Design

Might be used to estimate the effect of any phenotype on any outcome, if you can identify a gene that affects the phenotype and that gene has no other reason to be associated with the outcome.

For example, could be applied to estimate the health or cognitive effects of a:

1) Protein that is the direct product of a gene
2) Protein whose degradation or creation is under control of a gene
3) Behavior influenced by gene product
4) Affective state influenced by gene product
5) Health or cognitive state influenced by the gene product

Intuition for IV for Epidemiologists: RCT

- Randomization can be thought of as a special case of an instrumental variable:

Random assignment→X→Y

  - We wish to test whether X affects Y
  - We randomly assign people to receive treatment or exposure X
  - We compare the outcome Y across levels of randomization (ITT), rather than across levels of exposure or take-up
  - With imperfect compliance, we assume ITT is an underestimate of the causal effect of X on Y

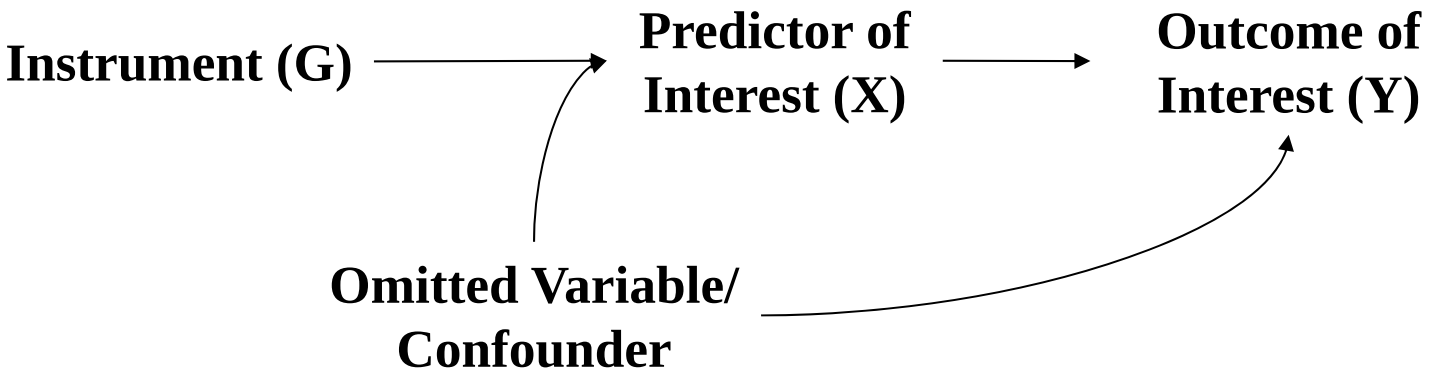
Intuition for IV

- The ITT is a valid test for the null hypothesis of no average causal effect of X on Y if:
  - There is at least some take-up (randomization affects exposure)
  - Randomization is fair (no common cause of randomization and the outcome)
  - Randomization influences the outcome *only* via the treatment X (not via related treatment X' or via compensatory pathways in the controls)
- These criteria for a valid RCT correspond exactly with the criteria for a valid IV/MR analysis

Instrument Assumptions: in intuitive terms

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

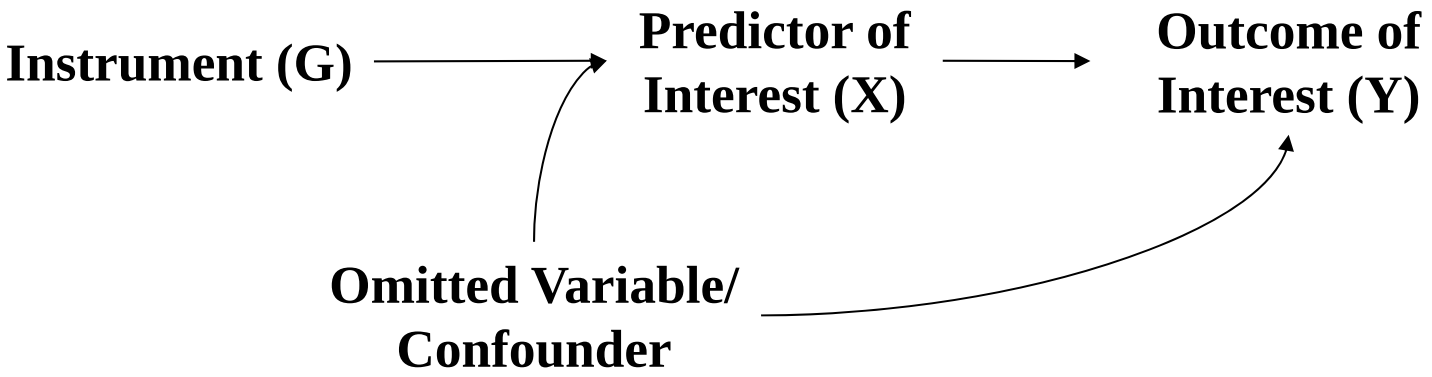
- G predicts X
- X does not affect G
- G has no effect on Y unless the effect is mediated by X
- No other variables influence both G and Y



Instrument Assumptions: in DAG Terminology

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

- G is not an effect of X
- G is not independent of X (typically: G affects X)
- Every unblocked path between G and Y contains an arrow pointing into X

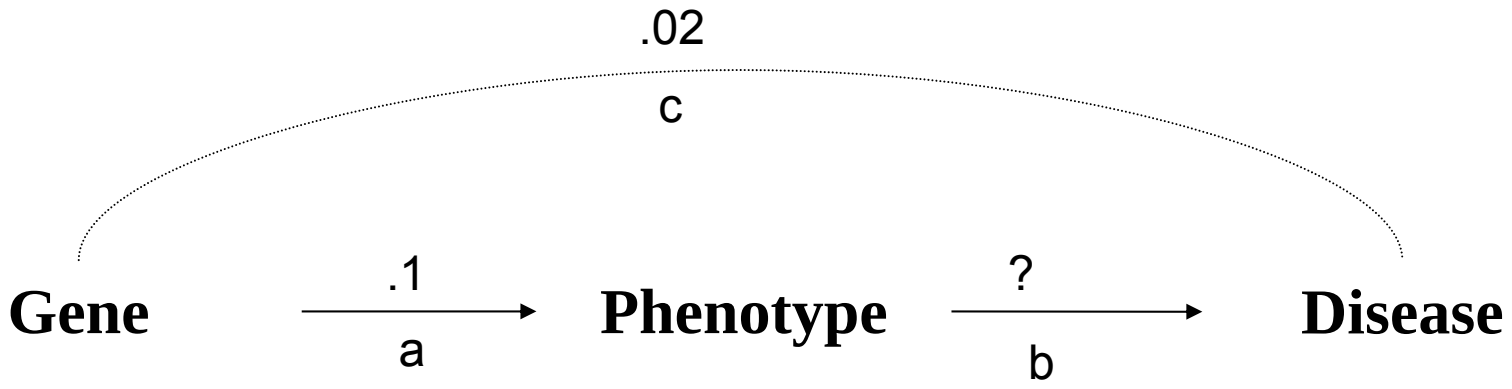


D-separation

- A path between two variables is blocked if:
  1. The path contains a non-collider that is in z, or
  2. The path contains a collider which is not in z, and no descendent of the collider is in z.

- If there is an unblocked path linking x and y, then x and y will typically be statistically dependent (unless there is a perfectly offsetting balance between two paths).

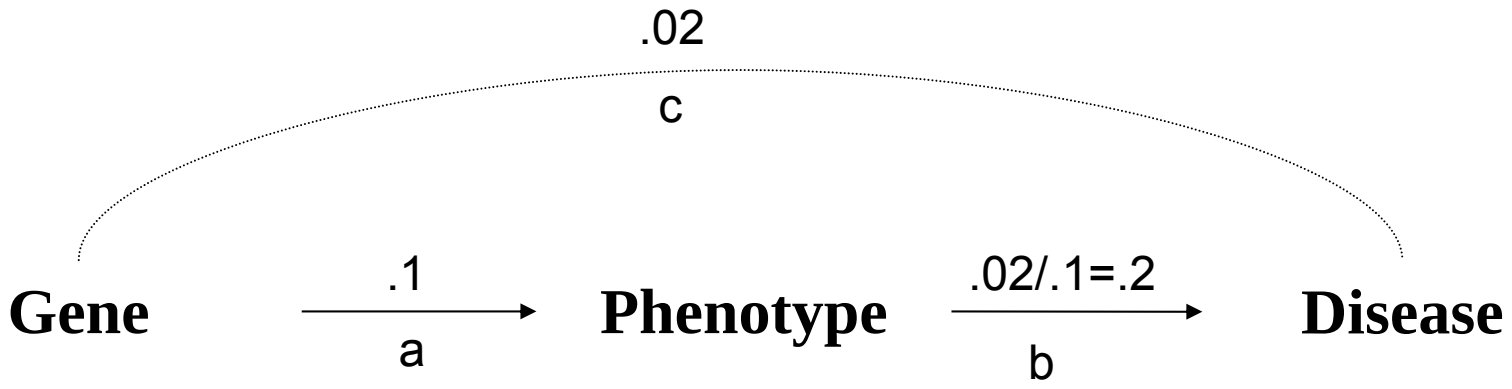
Obtaining IV Effect Estimates



Wald Estimates:

$$\text{IV Effect} = \frac{\Pr(\text{Disease}|\text{Gene}=1)-\Pr(\text{Disease}|\text{Gene}=0)}{\Pr(\text{Phenotype}|\text{Gene}=1)-\Pr(\text{Phenotype}|\text{Gene}=0)}$$

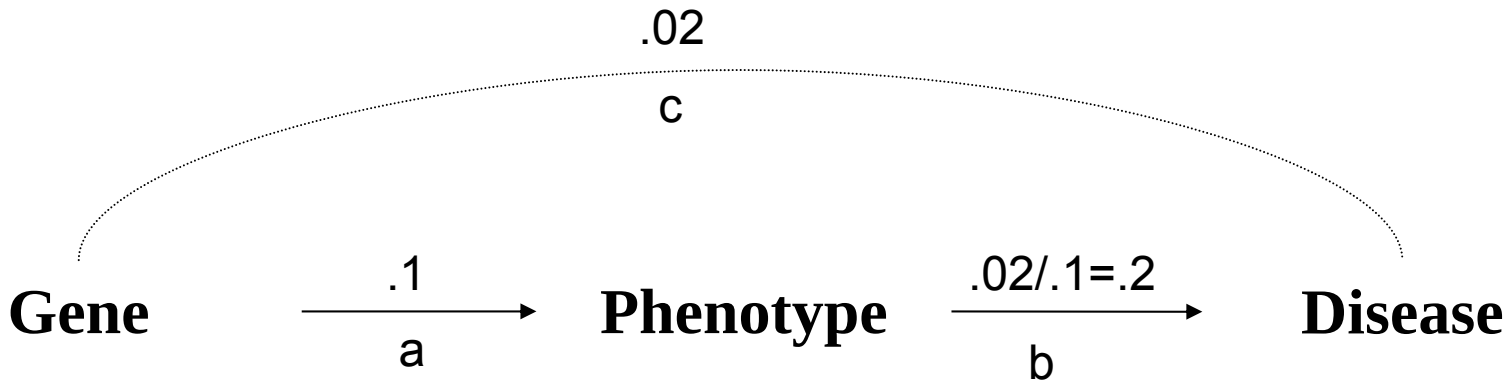
Obtaining IV Effect Estimates



Wald Estimates:

$$\text{IV Effect} = \frac{\Pr(\text{Disease}|\text{Gene}=1)-\Pr(\text{Disease}|\text{Gene}=0)}{\Pr(\text{Phenotype}|\text{Gene}=1)-\Pr(\text{Phenotype}|\text{Gene}=0)}$$

Obtaining IV Effect Estimates



Wald Estimates:

$$\text{IV Effect} = \frac{\Pr(\text{Disease}|\text{Gene}=1)-\Pr(\text{Disease}|\text{Gene}=0)}{\Pr(\text{Phenotype}|\text{Gene}=1)-\Pr(\text{Phenotype}|\text{Gene}=0)}$$

IV was developed with a focus on binary variables (instrument yes/no; phenotype present/absent; disease present/absent), so conventional estimator is based on risk differences but correlations are fine if you assume everything is linear: c=a*b and therefore b=c/a.

Variations on IV Estimators

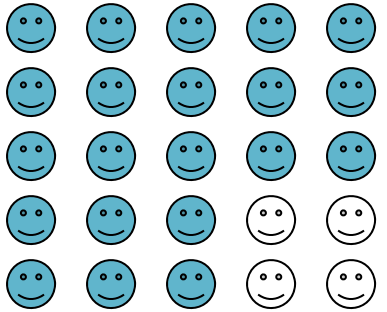
Wald Estimates:

Effect = $\frac{\text{(Difference in Avg Outcome between Genotypes)}}{\text{(Difference in Avg Phenotype between Genotypes)}}$

- Two Stage Least Squares Advantages (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
- Generalized method of moments
- Residual control approaches

Interpreting IV Effect Estimates

- Any given exposure may have different effects on different people in the population



- IV effect estimates are not the same as population average treatment effect (i.e., what would happen to the average value of Y if I treated everyone in the whole population).

- Alternative assumptions are required to support any specific interpretation of the IV estimate, the most common is "treatment effect on compliers".

Whose Causal Effect?

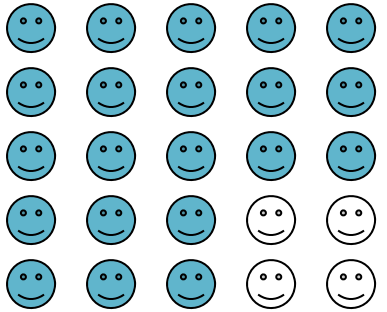
What the person will do if assigned to control treatment:

What the person will do if assigned to experimental treatment:

	Take control	Take experimental
Take control	Never-Takers	Compliers
Take experimental	Contrarians/ Defiers	Always Takers

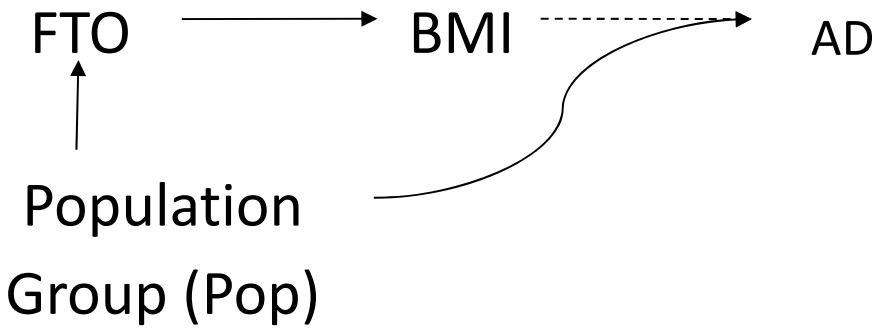
Interpreting IV Effect Estimates

- The effect of exposure (X) on the outcome (Y) among those people whose exposure was determined by the gene.
- This may or may not be the same as the effect of the exposure on *other* people, much less the whole population

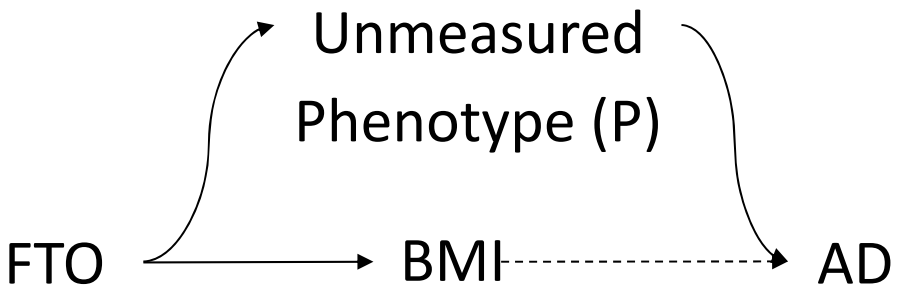


- It may not even be the same *sign* as the effect of the exposure on the whole population
- Impossible to know precisely whose exposure was determined by the gene

Causal Structures Violating IV Assumptions: Population stratification and Pleiotropy

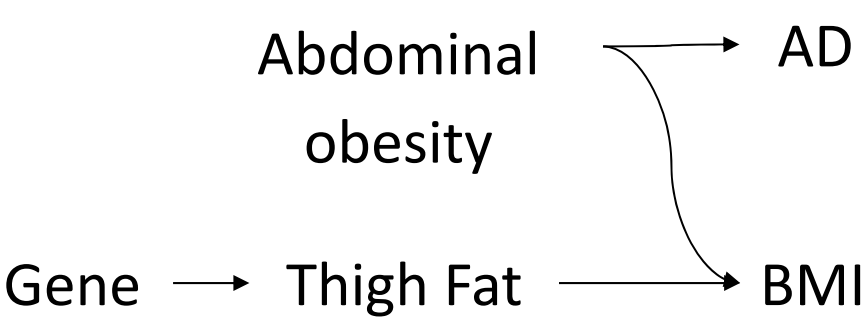


Population Stratification



Pleiotropy

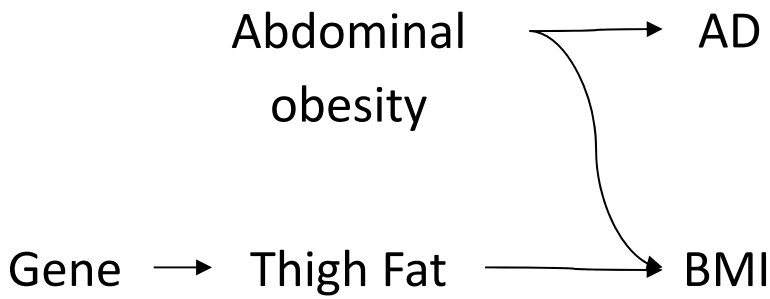
Causal Structures that (may) Violate IV Assumptions: multi-component phenotype



Phenotypes with multiple versions or components

- Gene predicts BMI
- No other pathways link the gene to AD
- A component of BMI (abdominal obesity) affects AD
- Gene will be independent of AD
- IV analysis will suggest the incorrect inference that BMI does not affect AD

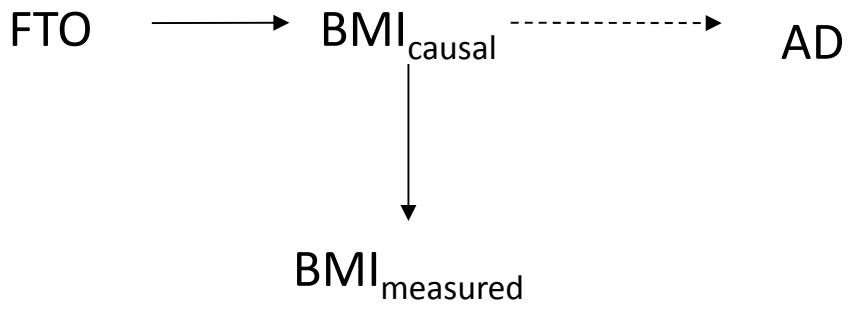
Causal Structures that (may) Violate IV Assumptions: multi-component phenotype



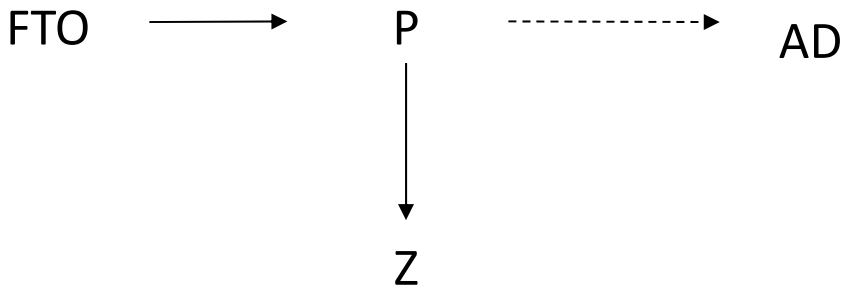
Phenotypes with multiple versions or components

- IV analysis will suggest the incorrect inference that BMI does not affect AD
- Correct inference: the phenotype affected by the gene (thigh fat) does not affect AD.
- In theory, could use this to help identify causally relevant variations on the phenotype (i.e. is it important to distinguish between abdominal or peripheral obesity? How about knee fat and ankle fat?)

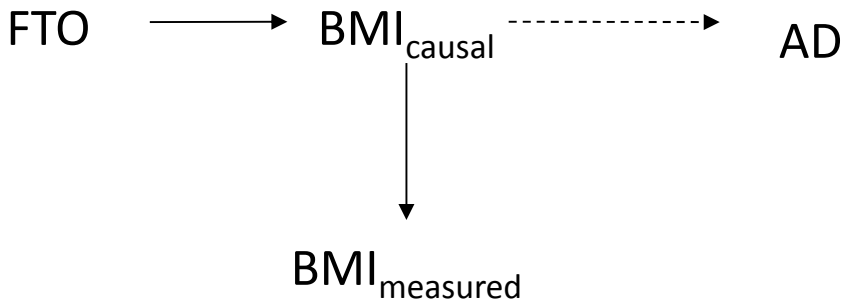
Violating IV Assumptions: mismeasuring the causal phenotype



Structurally related to "pleiotropy": IV estimate of the effect of ***Z*** on AD corresponds to effect of ***P*** on AD only under special circumstances



Mismeasuring the causal phenotype



This structure will almost always be relevant in MR, because the gene affects lifelong values of the phenotype, but we measure at only one or a few moments.

Appropriate to test the null hypothesis: BMI has no effect on AD. Therefore, helpful to present the "ITT" estimate of the association between the gene and the outcome.

IV estimate with $BMI_{measured}$ can be the same as, larger, or smaller than the IV estimate with BMI_{causal}

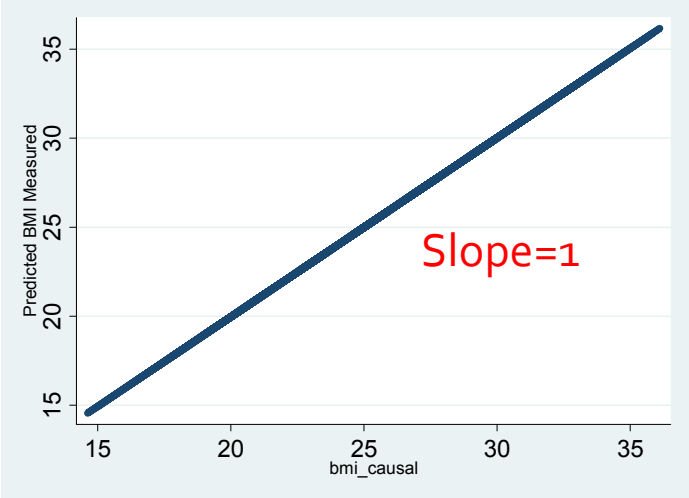
Mismeasuring the causal phenotype

How does $IV_{measured}$ relate to IV_{causal}?

FTO → BMI_{causal} ⇢ AD

↓

$BMI_{measured}$



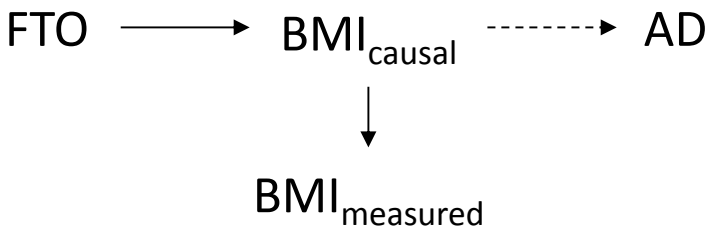
Classical measurement error:

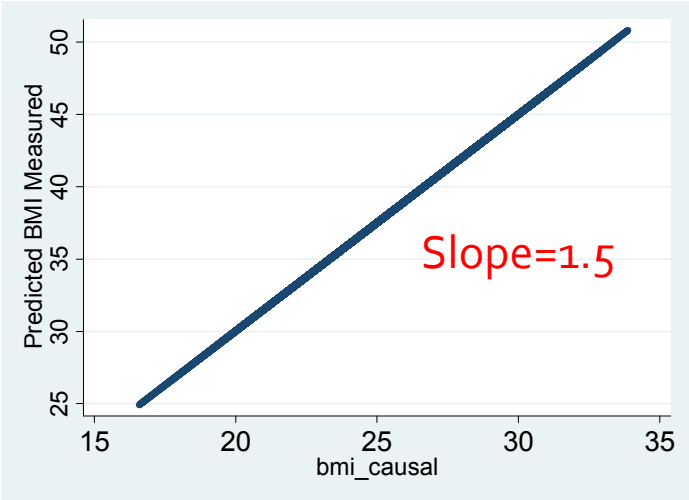
$BMI_{measured} = BMI_{causal} + e$

$IV_{measured} = IV_{causal}$

Mismeasuring the causal phenotype

How does $IV_{measured}$ relate to IV_{causal}?





$BMI_{measured}=1.5*BMI_{causal}+e$

If the IV estimate with BMI_{causal} >0 then:

$IV_{measured} < IV_{causal}$

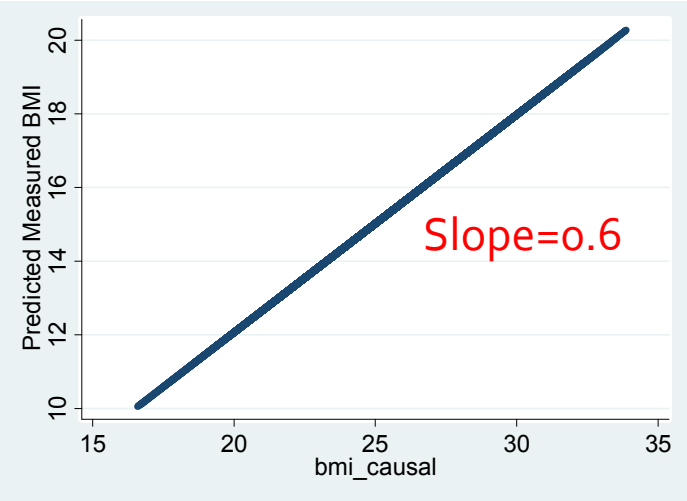
Mismeasuring the causal phenotype

How does $IV_{measured}$ relate to IV_{causal}?

FTO → BMI_{causal} ⇢ AD

↓

$BMI_{measured}$

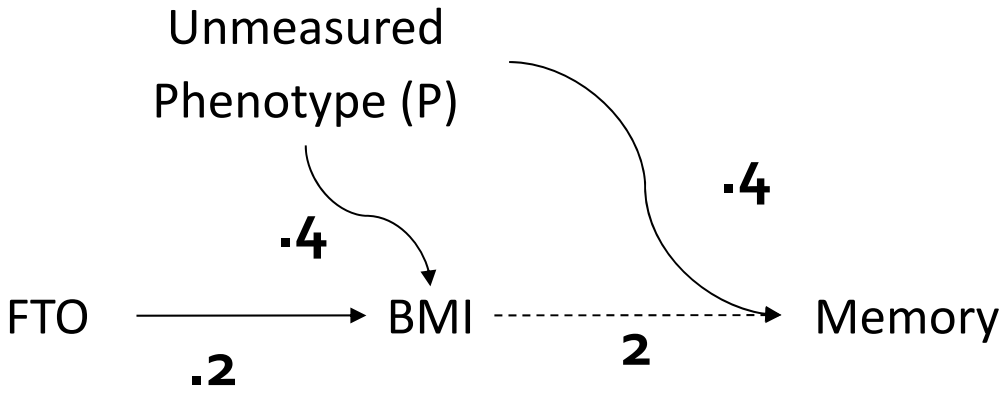


$BMI_{measured}=0.8*BMI_{causal}+e$

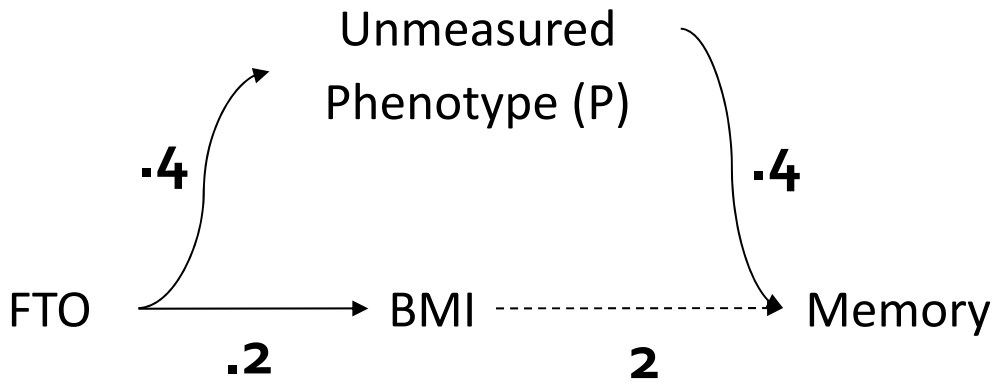
If the IV estimate with BMI_{causal} >0 then:

$IV_{measured} > IV_{causal}$

MR is sensitive to violations: pleiotropy

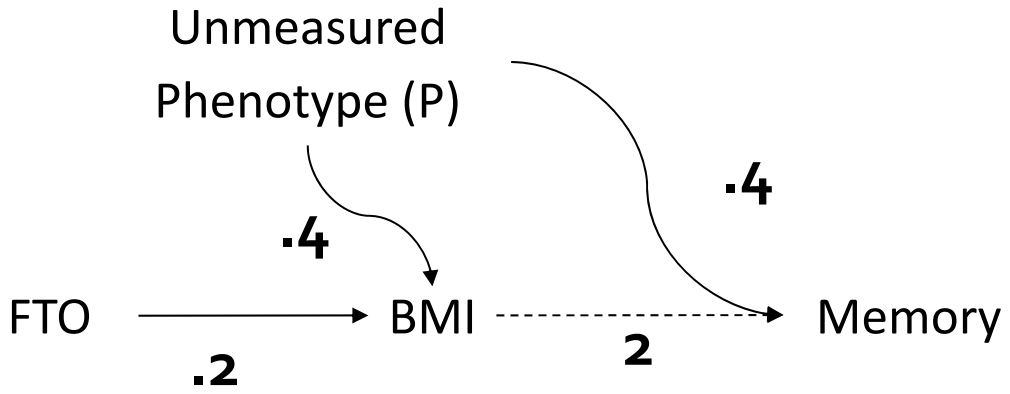


OLS estimate: 2.16

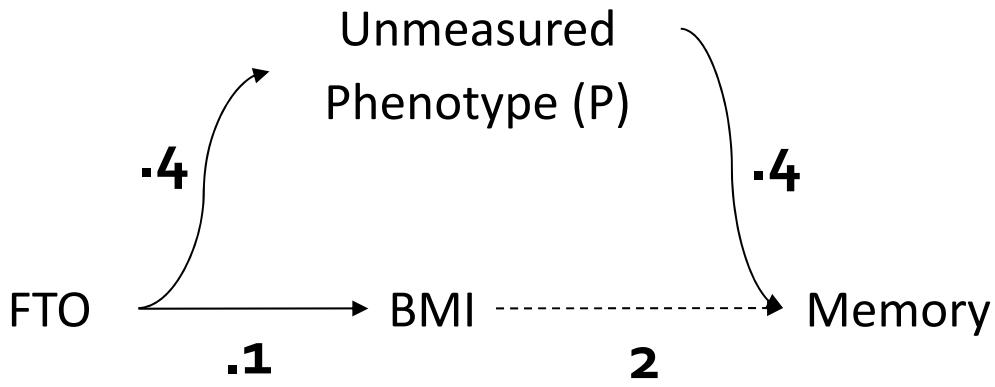


IV estimate: 2.8

MR is sensitive to violations: pleiotropy with a weaker instrument



OLS estimate: 2.16



IV estimate: 3.6

Improving MR Studies

- Stronger genetic determinants of the phenotype
  - Multiple genes: increasingly feasible in GWAS
  - Poly/multi-genic scores
- Sensitive tests for assumptions

Polygenic scores as first stage of MR

- Candidate gene approach
  - Allele count
  - Empirically weighted
- Z=PRS=b1*SNP1+b2*SNP2+b3*SNP3+bk*SNPk
- Where each b1…bk is based on the best available evidence on the effect of the SNP on the phenotype, e.g., from a recent large meta-analysis of GWAS studies.
- And each SNP1…SNPk is an allele count
- Under this formula, the value of the polygenic risk score is the predicted value if you specified a linear regression model with the best known betas.

Polygenic scores as first stage of MR

- Candidate gene approach
  - Allele count
  - Empirically weighted
- Z=PRS=b1*SNP1+b2*SNP2+b3*SNP3+bk*SNPk
- Where each b1…bk is based on the best available evidence on the effect of the SNP on the phenotype, e.g., from a recent large meta-analysis of GWAS studies.
- This is appropriate only if you assume your sample corresponds with the population in the prior meta-analysis.
- Improves statistical power if meta-analyzed betas are better estimates than internal betas.
- Eliminates weak instruments bias.

Polygenic scores as first stage of MR

- Candidate gene approach
  - Allele count
  - Empirically weighted
- Genome wide scoring
  - We don't care about the G→X mechanism, just need to predict X
  - Genome wide scores (e.g. all p<.05) might work
- Limitation: any new gene can violate the IV assumptions
- Small violations can lead to big biases

Evaluating the assumptions

1. Constraints implied by theory
2. Over-identification tests
3. IV inequality constraints
4. Stratification-based tests

1 Evaluating MR Studies: leverage prior assumptions about confounding

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

2 Evaluating MR Studies: Over-identification tests

- Use multiple instrumental variables to conduct over-identification tests.
- Other genes or even polymorphisms of the same gene might provide additional instruments.
- 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.

3 Evaluating MR Studies: instrumental inequality tests

- These tests are applicable only when the causal phenotype is known to be categorical.
- 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 MR Studies: Modifying factors

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

Conclusions

- Host of research questions potentially amenable to MR
- Cheap, "easy" (!!!), applicable to stubborn questions that are expensive and hard to study with trials
- Assumptions are strong, often seem implausible, so efforts to test are critical for credible MR
- GW data: more data, more opportunities to answer the question (more opportunities to incorrectly answer the question?)
- Routinely implement assessments of instrument validity
- Major challenge: statistical power because genes aren't very strong predictors

Acknowledgments

- NIMH (RC4MH092707) MR studies in GWAS
- Jamie Robins
- Eric Tchetgen Tchetgen
- Stefan Walter
- David Rehkopf
- Ezra Susser
- Bruce Link
- Laura Kubzansky