Some topics from Mendelian randomization

Tom Palmer, Lecturer in Medical Statistics, Department of Mathematics and Statistics
tom.palmer@Lancaster.ac.uk

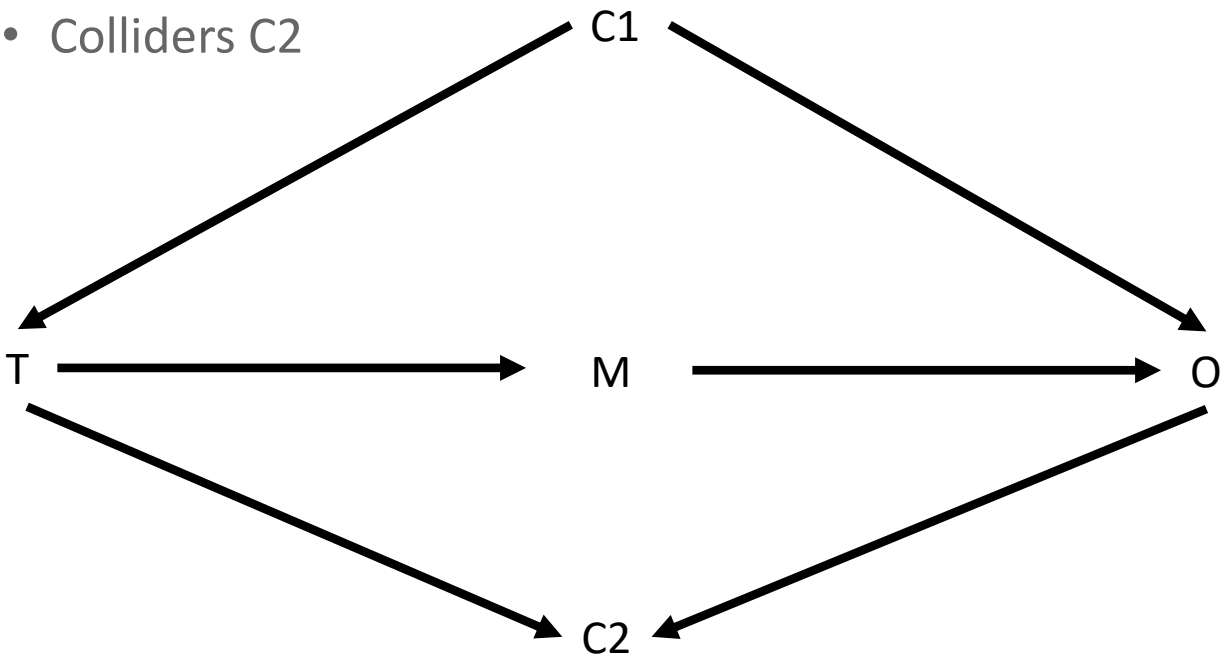
Outline

- Directed Acyclic Graphs for statistical modelling
- Mendelian randomization basics
  - Assumptions
  - Risk difference example
  - Risk ratio example
- Summary

Directed Acyclic Graphs for statistical modelling

Lancaster University

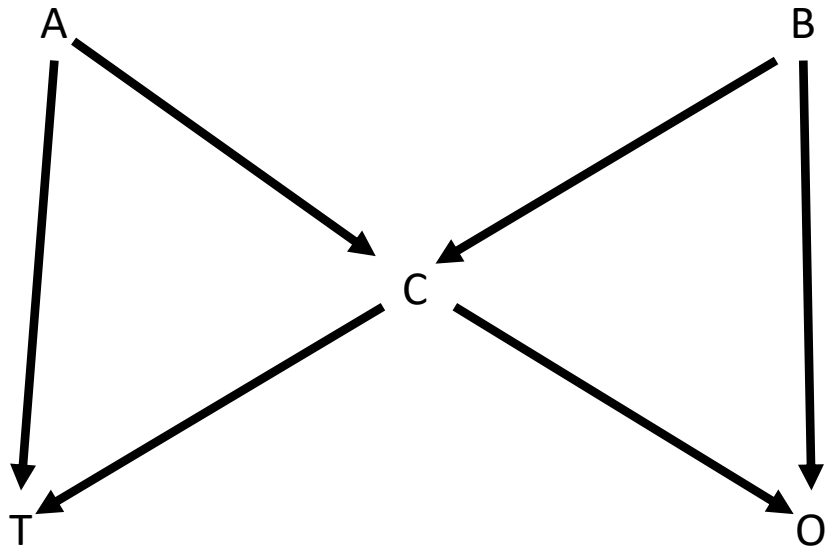
- As long as the DAG is not cyclic the rules of conditional independence hold
- Advantages of DAGs:
  - Good at depicting:
    - Confounding C1
    - Mediation M
    - Colliders C2

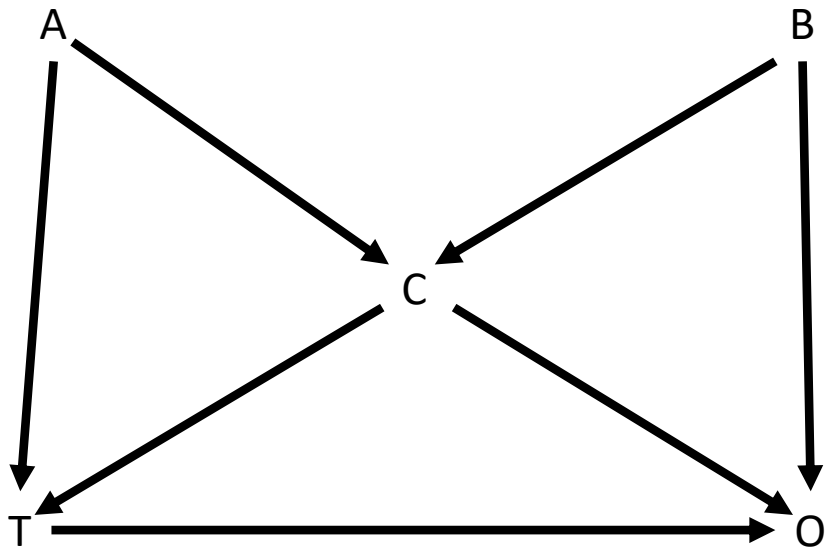


– Good at depicting conditional independence

- Regress O on T and adjust for M
  - if the effect of T does not go to null then you can argue there must be another pathway between T and O

- Given the correct model they can tell us when we have adjusted for “enough” variables.
  - In the terminology of DAGs we must block all backdoor paths between the Treatment and the Outcome

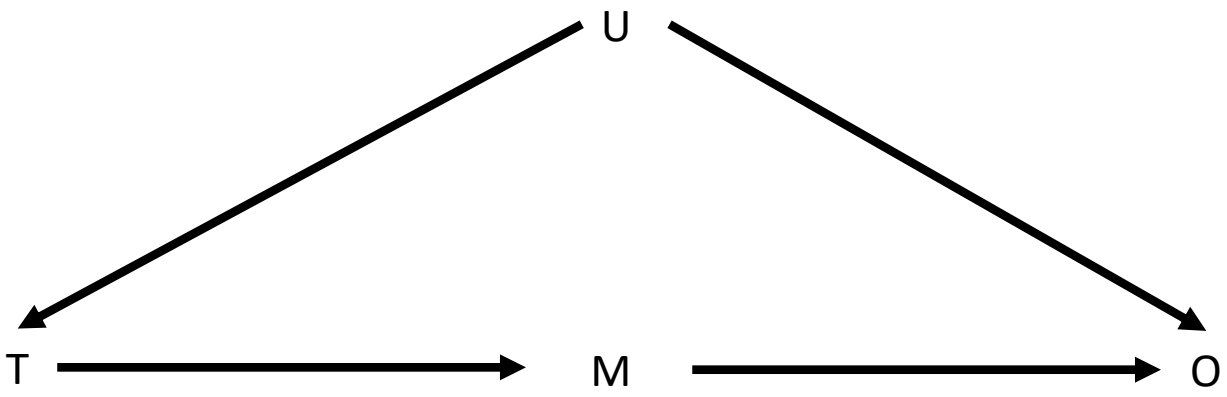




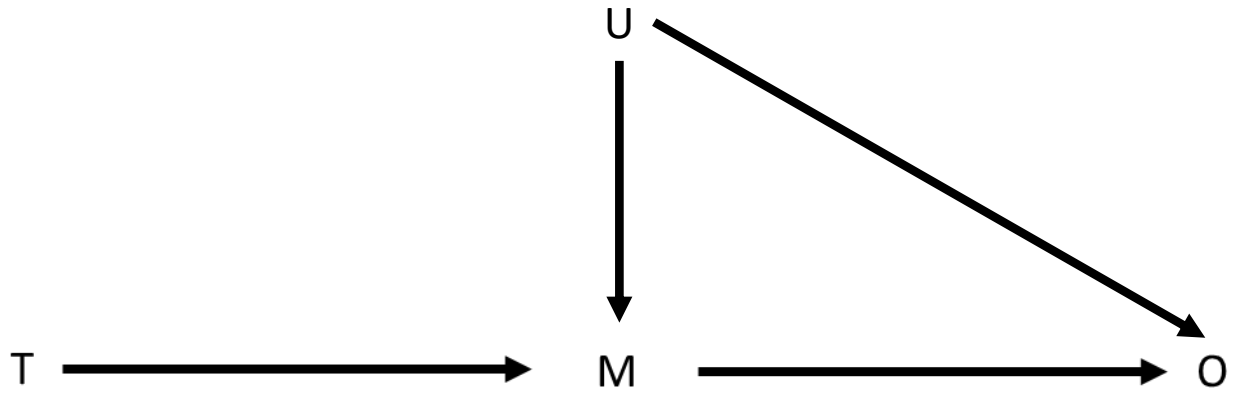
- To estimate the effect of T on O what do we need to adjust for?
- What backdoor paths are there?
  - Starts with an arrow going into T; then arrows can go in either direction
  - Block them by adjusting for variables on them
  - Watch out for induced collider bias

Answer: C and B; or C and B; or A, B, and C.

- They can tell us when certain indirect estimates are very useful



- Imagining unmeasured confounders can tell us the potential weaknesses in our models:



- What happens if we adjust for the mediator?

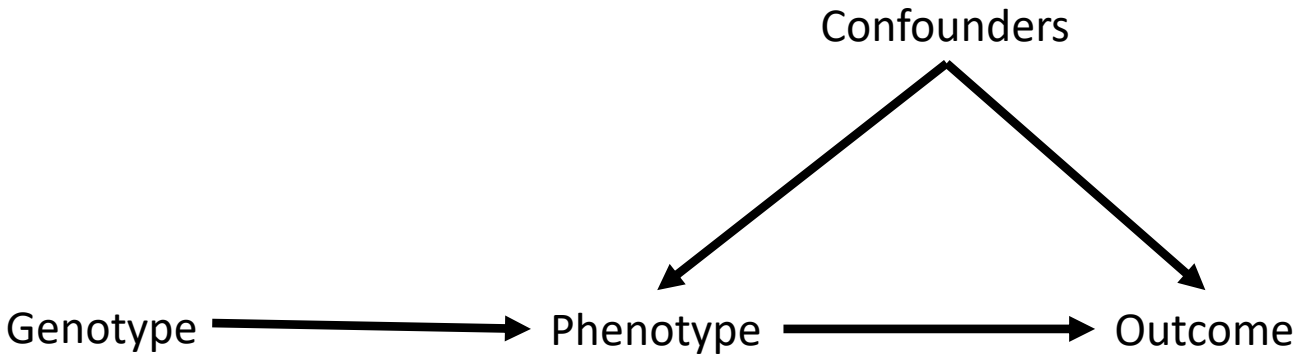
- Essentially provide a formal mathematical framework for the old statistical modelling guidelines:
  - Adjust for confounders
  - Don’t adjust for something on the causal pathway (unless you want to partition the effect into its direct/indirect components)
  - Don’t adjust for a consequence of the outcome

- Realistically complex framework of what to adjust for, neither of:
    - brought about a 10% change in treatment effect
    - was statistically significant in the model (but what if not an confounder or independent predictor)

- Disadvantages of DAGs:
  - Don’t telling us how big the impact of confounding/collider bias will be (will it actually affect our analysis in a meaningful way)
  - Difficult to represent effect modification (interaction) – although some proposals
  - Don’t tell us about other structures, e.g. random effects

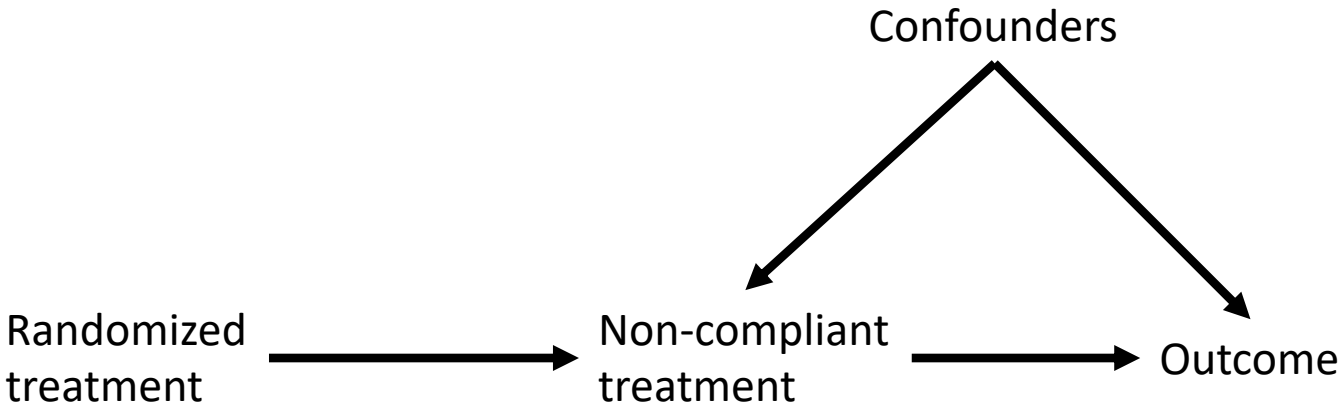
Mendelian randomization

- Davey Smith (2003) realised that genotypes could be used as instrumental variables in epidemiological studies



- Genotype associated with Phenotype
    - robustly, i.e. previous GWAS
- Genotype only affects Outcome through Phenotype
    - Exclusion restriction – can be hard to justify
- Genotype independent of all measured and unmeasured confounders
    - The **randomization**; Gregor Mendel's second law
- **Can't test 2 and 3 fully with observational data**

- Instrumental variables have been used in several different study types
- Clinical trials

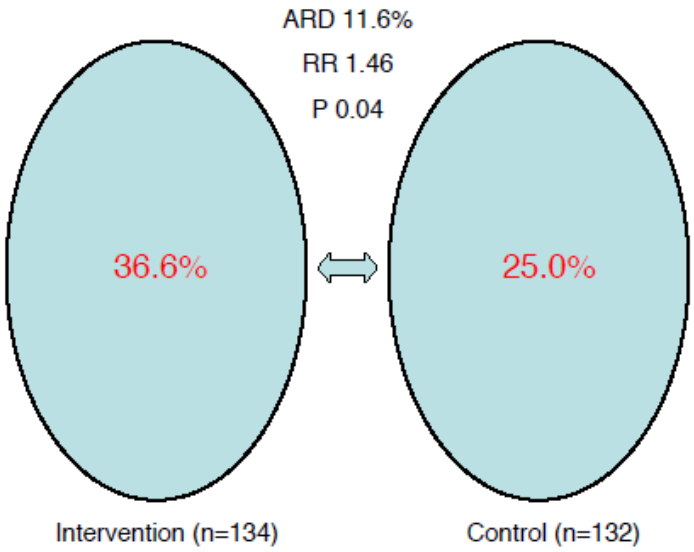


- Randomized variables can occur in economics etc., e.g. draft lotteries for Vietnam war

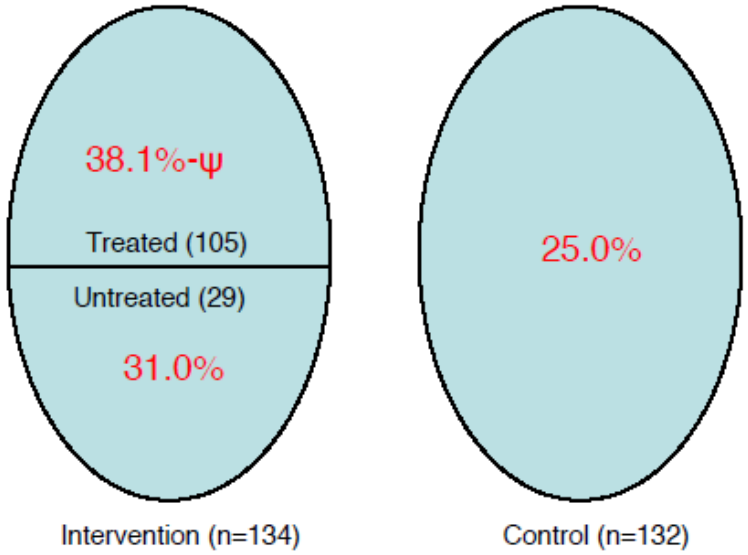
- With individual level data many IV estimators
  - Continuous outcome
    - Two-stage least squares
    - Two-stage residual inclusion estimators
  - Binary outcome
    - Two-stage residual inclusion estimators
    - Structural mean models

Linear IV / additive structural mean model example

- Tenhave et al., JASA, 2004
- 266 African American adults with high cholesterol and/or hypertension
- Control group: usual care (nutritional information)
- Intervention: usual care plus audio tapes
- Outcome: beneficial change in cholesterol
- Naïve analysis



- However there was non-compliance in the intervention group

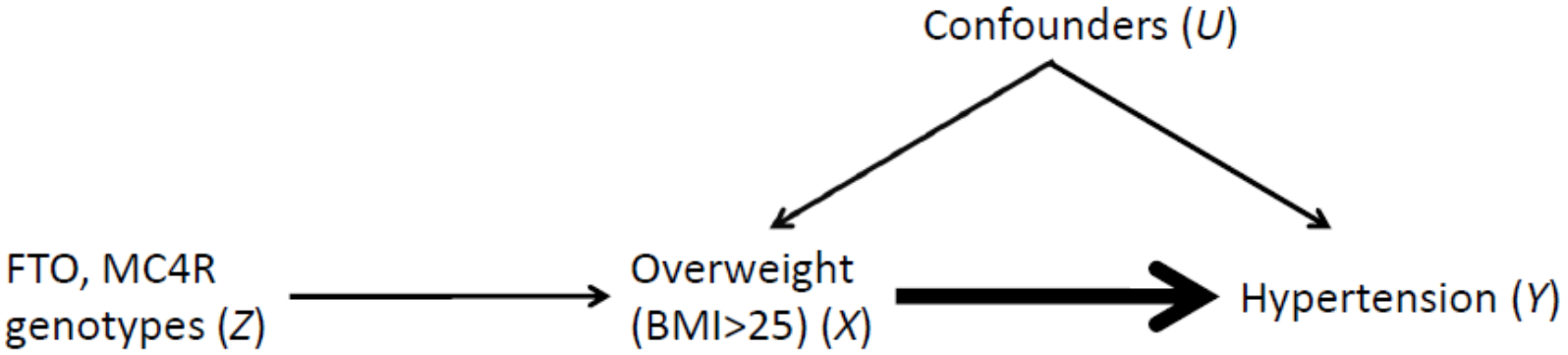


- IV ratio $$\psi = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(X|Z=1) - E(X|Z=0)} = \frac{36.6 - 25.0}{105/134 - 0}$$
$$= 11.6/78.4 = 14.8\% \text{ (95\%CI } 0.8\%, 28.7\%; P = 0.04)$$

G-estimation: what would have happened if no-one was treated

ASMM estimate: $(38.1 - \psi)(105/134) + 31.0(29/134) = 25.0$

$$\psi = 14.8\%$$



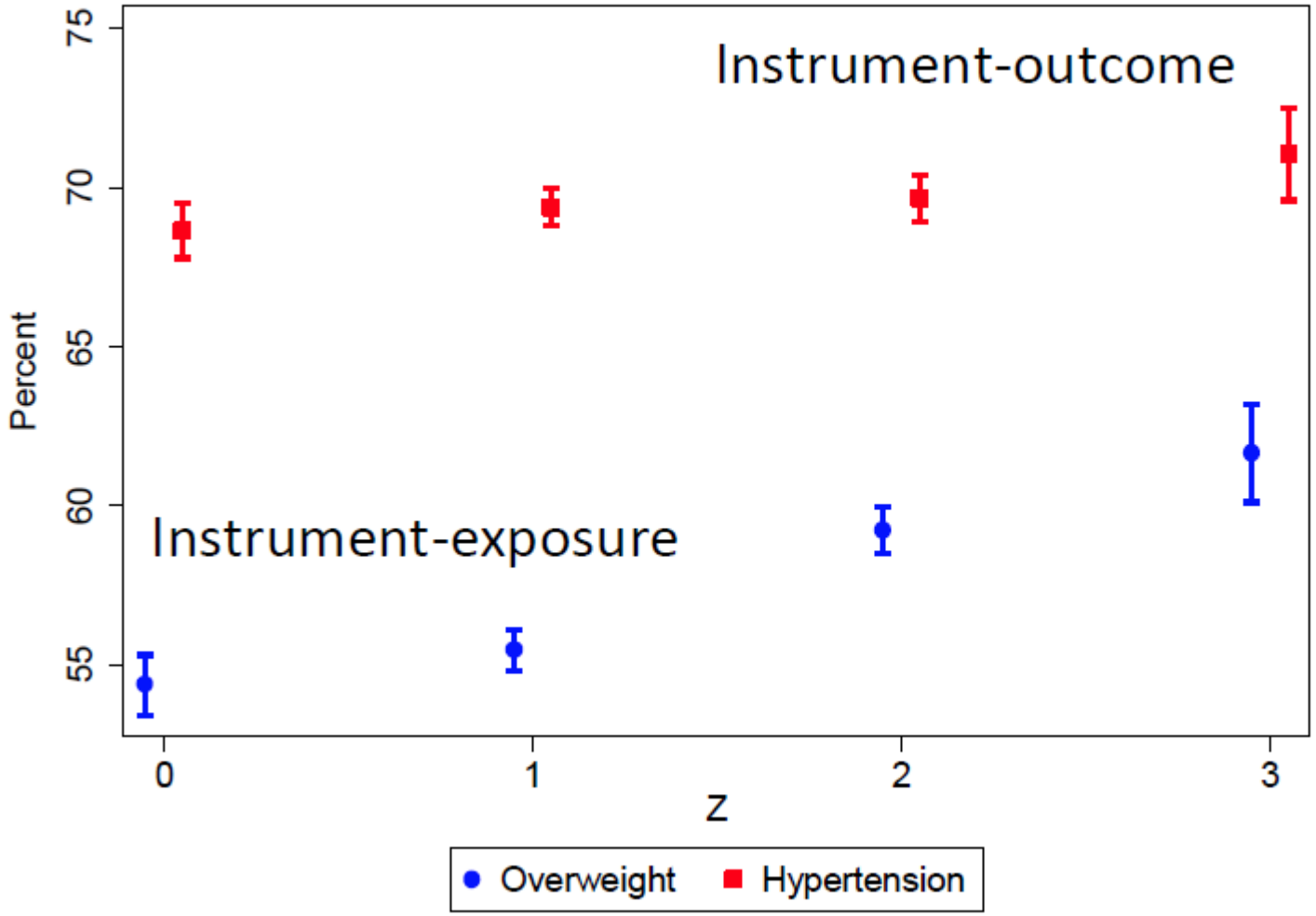
- 'Observational' association between overweight and hypertension

	No Hypertension	Hypertension	Total
Not Overweight	10,066 42%	13,909 58%	23,975
Overweight	6,906 22%	24,642 78%	31,548
Total	16,972 31%	38,551 69%	55,523 χ^2 P<0.001

- Risk ratio for hypertension 1.35 (1.32, 1.37)

Lancaster University

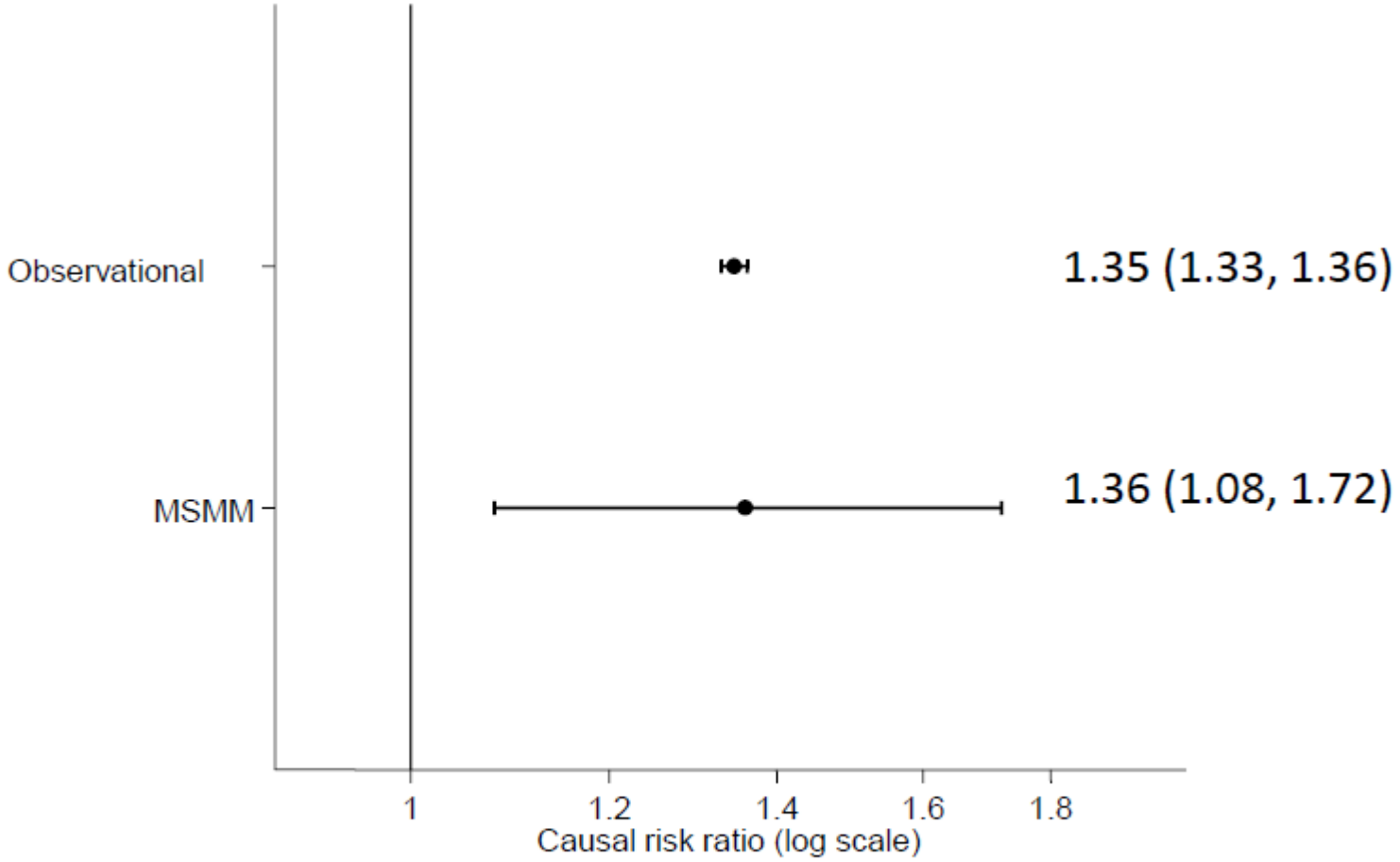
Exposure (over-weight) & outcome (hypertension) by instrument



P=0.007

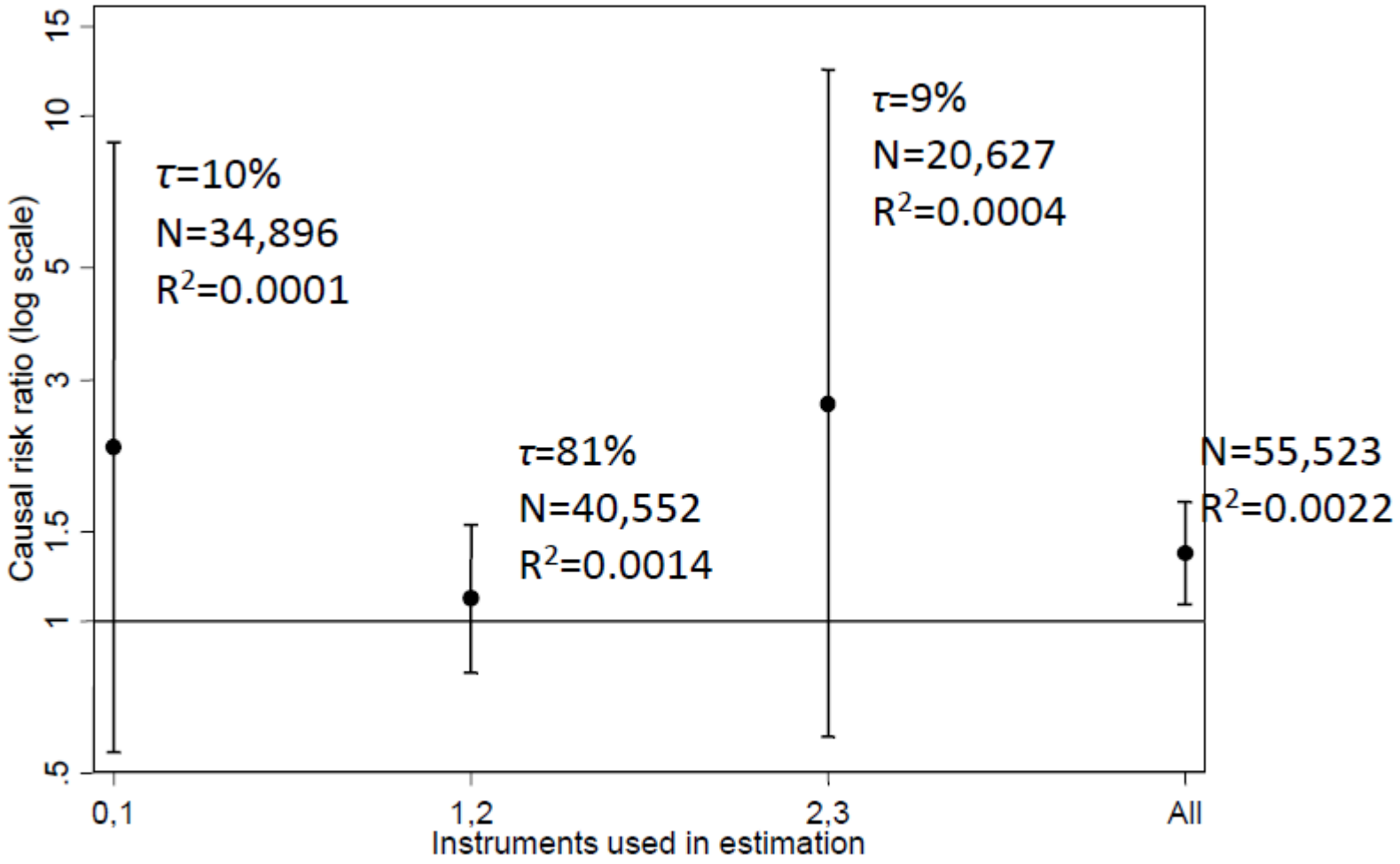
P<0.001

R^2=0.002



MSMM: Hansen over-identification test $P = 0.31$

$\mathrm{E}[Y(0)] = 0.58$ (0.50, 0.65)



Check: $(0.10 \times 2.21) + (0.81 \times 1.11) + (0.09 \times 2.69) = 1.36$

Summary

Lancaster University

- DAGs provide a realistically complex way of viewing statistical models
- Strengths – they can tell us what to adjust for
- Weaknesses – not good at showing effect modification
- In observational epidemiology genotypes can be used as instrumental variables
- Allow estimation of causal effects of phenotypes upon disease
- Important differences between estimates from a clinical trial:
  - Cohort studies usually contain wider age of people; and less strict entry criteria
- Estimation of different parameters with individual level data possible
- Recent developments (MR-Egger) use summary data