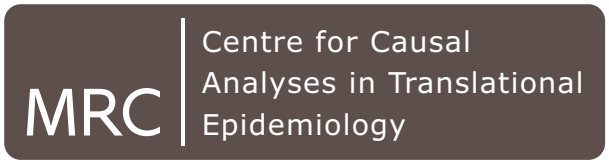
Including multiple instrumental variables in Mendelian randomization analyses

Tom Palmer Debbie Lawlor Jonathan Sterne

MRC Centre for Causal Analyses in Translational Epidemiology,
Department of Social Medicine, University of Bristol

26 August 2009



Outline

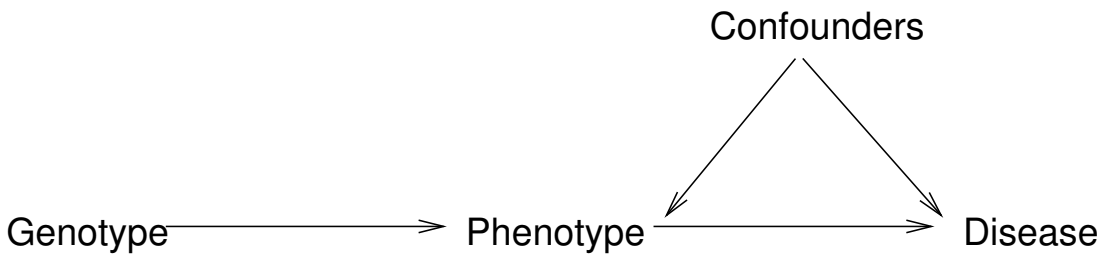
- Introduction to Mendelian randomization
- Multiple instruments example using ALSPAC data:
  - instrument strength
  - over-identification
  - allele scores
- Multiple instruments discussion

Introduction

Mendelian randomization approach:

- Difficult to adjust for all possible confounders
- Genotypes - instrumental variables
- Infer causal phenotype-disease association

(Davey Smith & Ebrahim, 2003)

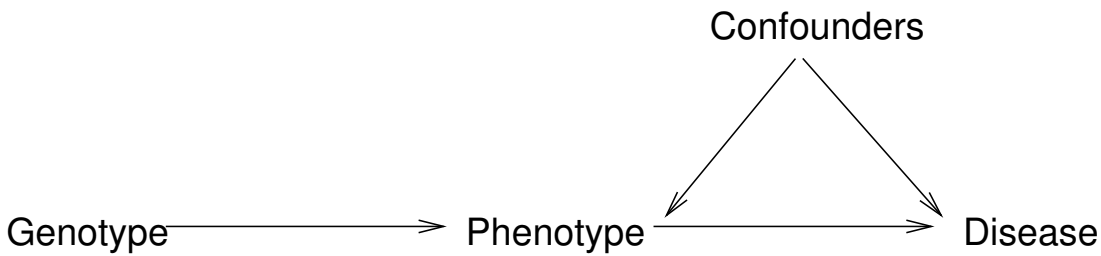


Introduction

Mendelian randomization approach:

- Difficult to adjust for all possible confounders
- Genotypes - instrumental variables
- Infer causal phenotype-disease association

(Davey Smith & Ebrahim, 2003)



IV assumptions, genotype should be:

(i) independent of confounders

(ii) associated with phenotype

(iii) independent of disease given phenotype and confounders

Problem:

- MR analyses have low power:
  - Weak instruments - bias IV estimate & wide CI
  - Genotypes explain small proportion of variability in phenotypes - small R^2 & wide CIs

Problem:

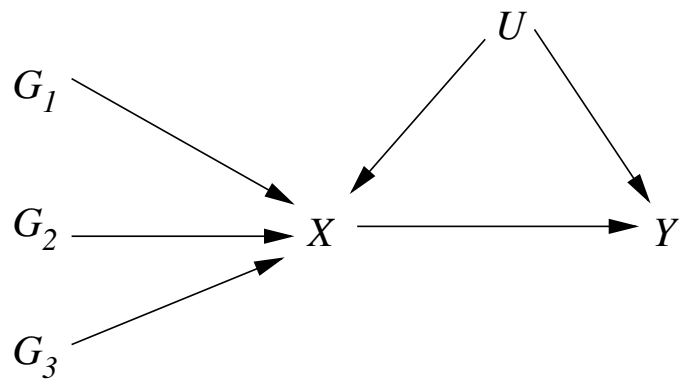
- MR analyses have low power:
  - Weak instruments - bias IV estimate & wide CI
  - Genotypes explain small proportion of variability in phenotypes - small R^2 & wide CIs

Solutions:

- Increase study sample size
- Stronger instrument
- **Multiple instruments**
- (Meta-analysis)

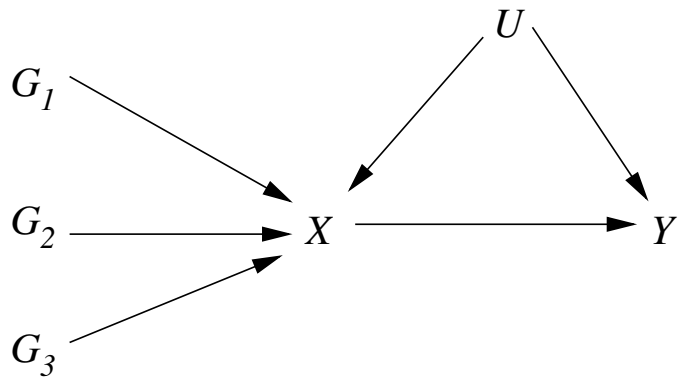
Multiple instruments

- Ideal situation (Didelez & Sheehan, 2007):



Multiple instruments

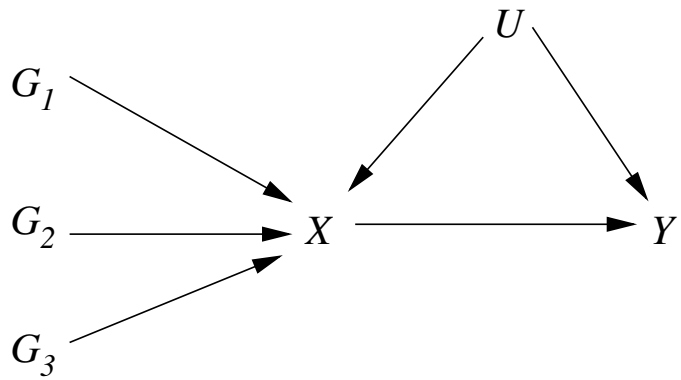
- Ideal situation (Didelez & Sheehan, 2007):



- 1 instrument: $0.015 < p < 0.05$ GP coefficient stat sig & weak (Lawlor et al., 2008)

Multiple instruments

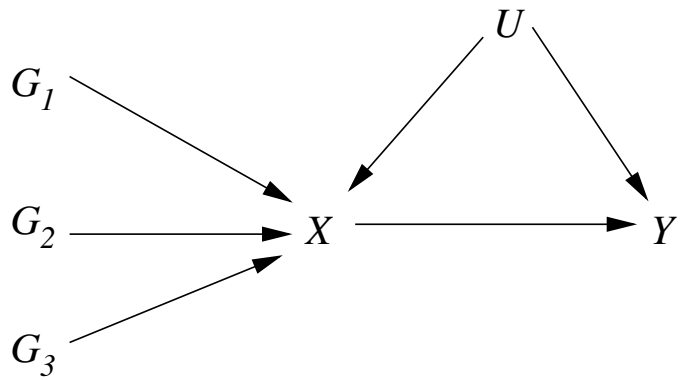
- Ideal situation (Didelez & Sheehan, 2007):



- 1 instrument: $0.015 < p < 0.05$ GP coefficient stat sig & weak (Lawlor et al., 2008)
- Multiple instruments: Cragg-Donald F-statistic (Cragg & Donald, 1993; Stock et al., 2002)

Multiple instruments

- Ideal situation (Didelez & Sheehan, 2007):



- 1 instrument: $0.015 < p < 0.05$ GP coefficient stat sig & weak (Lawlor et al., 2008)
- Multiple instruments: Cragg-Donald F-statistic (Cragg & Donald, 1993; Stock et al., 2002)
- Over-identification: Sargan/Hansen test

Is fat mass causally related to bone mineral density?

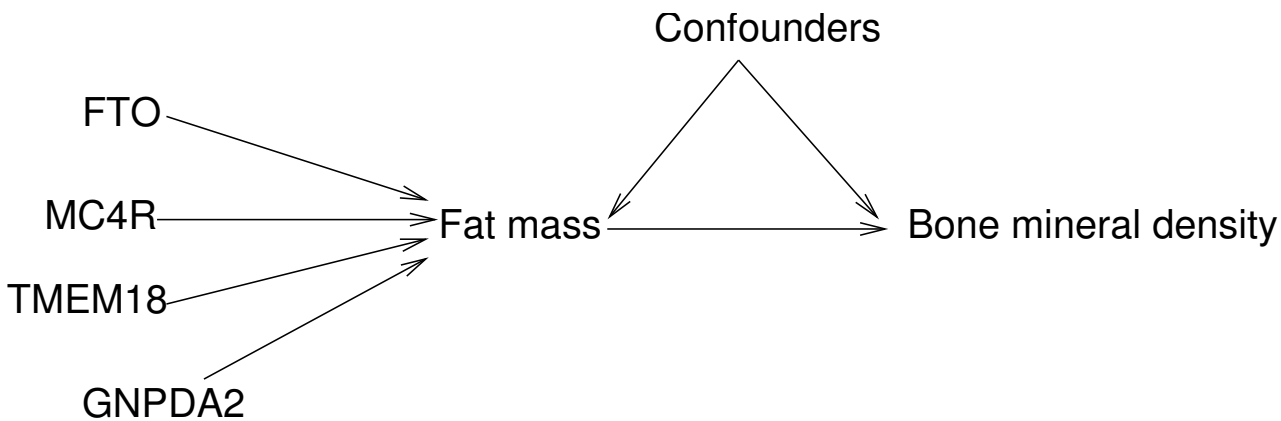
- Outcome: bone mineral density
- Phenotype: fat mass (DXA scan)

Is fat mass causally related to bone mineral density?

- Outcome: bone mineral density
- Phenotype: fat mass (DXA scan)
- IVs: FTO, MC4R, TMEM18, GNPDA2
  chromosomes 16, 18, 2, 4
  (Frayling et al., 2007; Loos et al., 2008; Willer et al., 2009)

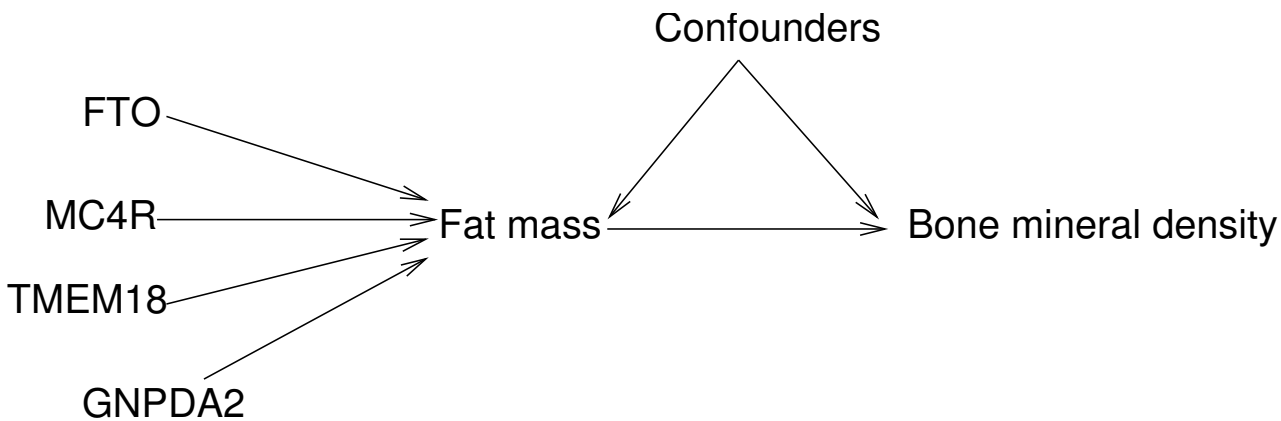
Is fat mass causally related to bone mineral density?

- Outcome: bone mineral density
- Phenotype: fat mass (DXA scan)
- IVs: FTO, MC4R, TMEM18, GNPDA2
  chromosomes 16, 18, 2, 4
  (Frayling et al., 2007; Loos et al., 2008; Willer et al., 2009)



Is fat mass causally related to bone mineral density?

- Outcome: bone mineral density
- Phenotype: fat mass (DXA scan)
- IVs: FTO, MC4R, TMEM18, GNPDA2
  chromosomes 16, 18, 2, 4
  (Frayling et al., 2007; Loos et al., 2008; Willer et al., 2009)



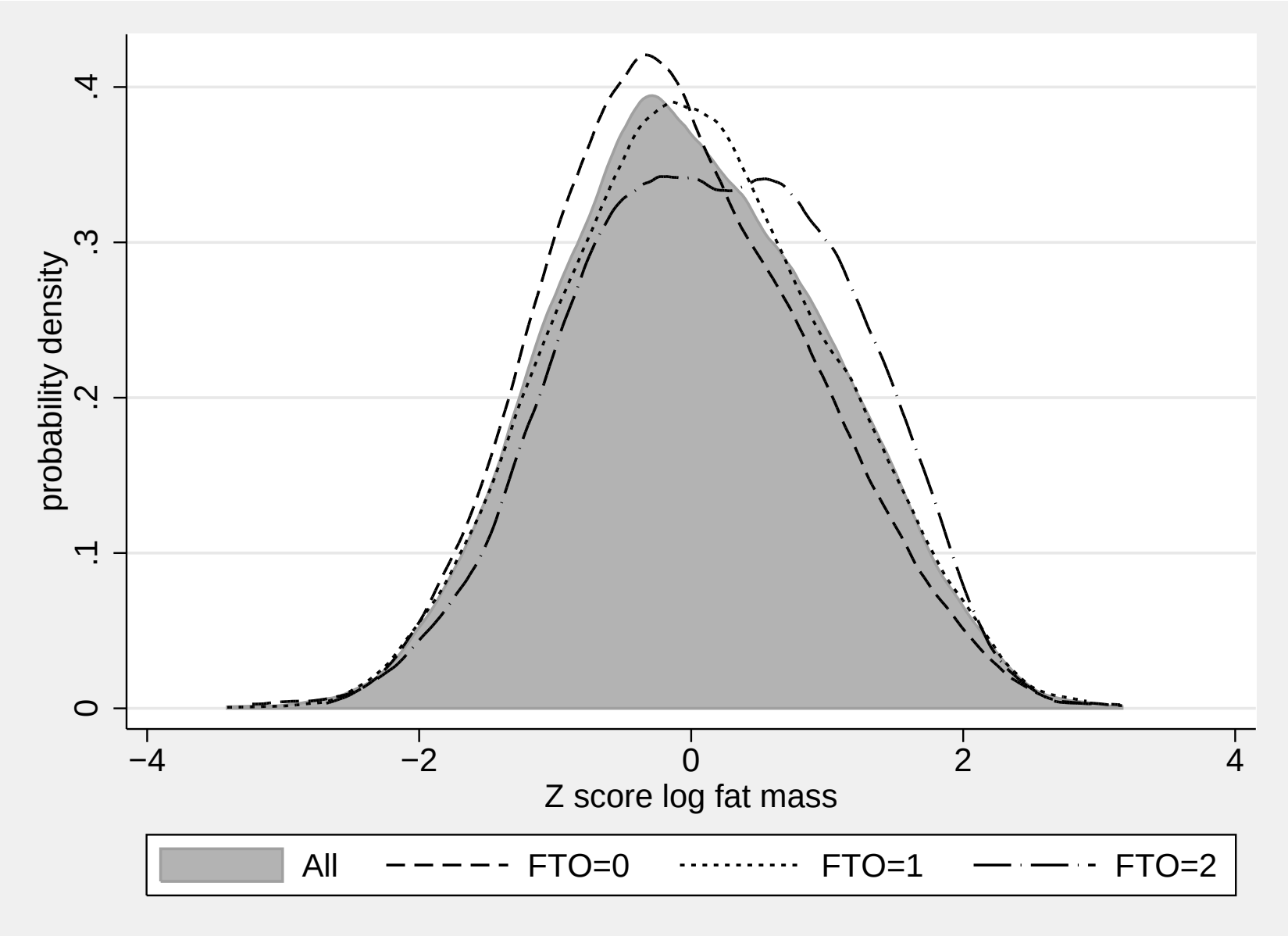
- FTO & MC4R: 0.2-0.4 kg/m^2 inc BMI
  OR: 1.1-1.3 for obesity (BMI $>$ 30 kg/m^2)

- Eligible sample: 5509 children, age 7-9yrs
- Fat mass & bone mineral density:
  - positively skewed - log & z-score
  - exponentiated regression coefficients - ratios of geometric means

- Eligible sample: 5509 children, age 7-9yrs
- Fat mass & bone mineral density:
  - positively skewed - log & z-score
  - exponentiated regression coefficients - ratios of geometric means
- Approaches:
  - IV with each SNP separately
  - IV with all four SNPs
  - IV with allele score - sum of risk alleles for each subject (Weedon et al., 2008)

- Eligible sample: 5509 children, age 7-9yrs
- Fat mass & bone mineral density:
  - positively skewed - log & z-score
  - exponentiated regression coefficients - ratios of geometric means
- Approaches:
  - IV with each SNP separately
  - IV with all four SNPs
  - IV with allele score - sum of risk alleles for each subject (Weedon et al., 2008)
- Estimation:
  - TSLS
  - AR/LIML, LM, CLR (Mikusheva & Poi, 2006)

CDFs of BMD by FTO genotypes



Model	Coef (95% CI)	P	F	R_p^2	DWH	Sargan	N
OLS	1.22 (1.19, 1.26)	< 0.001					4796

Model	Coef (95% CI)	P	F	R_p^2	DWH	Sargan	N
OLS	1.22 (1.19, 1.26)	< 0.001					4796
FTO	1.41 (1.05, 1.89)	0.023	45.4	0.01	0.32	NA	5091
MC4R	2.42 (1.42, 4.12)	0.001	20.0	0.004	0.002	NA	5412

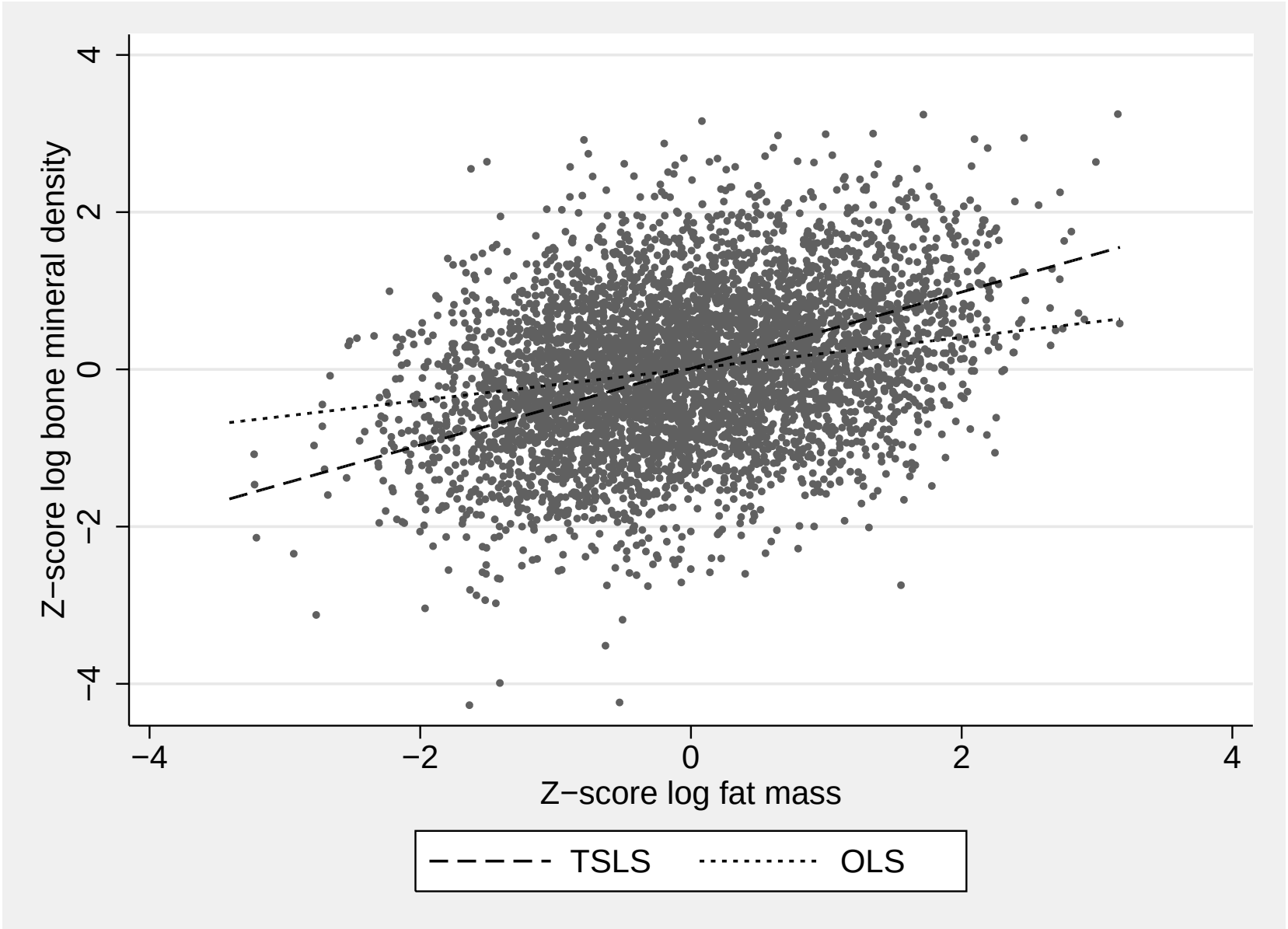
Model	Coef (95% CI)	P	F	R_p^2	DWH	Sargan	N
OLS	1.22 (1.19, 1.26)	< 0.001					4796
FTO	1.41 (1.05, 1.89)	0.023	45.4	0.01	0.32	NA	5091
MC4R	2.42 (1.42, 4.12)	0.001	20.0	0.004	0.002	NA	5412
TMEM18	2.17 (0.92, 5.12)	0.08	7.0	0.0013	0.13	NA	5323
GNPDA2	0.92 (0.42, 2.01)	0.84	6.9	0.0013	0.46	NA	5303

Model	Coef (95% CI)	P	F	R_p^2	DWH	Sargan	N
OLS	1.22 (1.19, 1.26)	< 0.001					4796
FTO	1.41 (1.05, 1.89)	0.023	45.4	0.01	0.32	NA	5091
MC4R	2.42 (1.42, 4.12)	0.001	20.0	0.004	0.002	NA	5412
TMEM18	2.17 (0.92, 5.12)	0.08	7.0	0.0013	0.13	NA	5323
GNPDA2	0.92 (0.42, 2.01)	0.84	6.9	0.0013	0.46	NA	5303
4 SNPs	1.63 (1.28, 2.06)	< 0.001	$18.6_{16.9}$	0.015	0.013	0.16	4796
AR/LIML	1.66 (1.29, 2.23)						
LM	(1.30, 2.21)						
CLR	(1.30, 2.20)						

Model	Coef (95% CI)	P	F	R_p^2	DWH	Sargan	N
OLS	1.22 (1.19, 1.26)	< 0.001					4796
FTO	1.41 (1.05, 1.89)	0.023	45.4	0.01	0.32	NA	5091
MC4R	2.42 (1.42, 4.12)	0.001	20.0	0.004	0.002	NA	5412
TMEM18	2.17 (0.92, 5.12)	0.08	7.0	0.0013	0.13	NA	5323
GNPDA2	0.92 (0.42, 2.01)	0.84	6.9	0.0013	0.46	NA	5303
4 SNPs	1.63 (1.28, 2.06)	< 0.001	$18.6_{16.9}$	0.015	0.013	0.16	4796
AR/LIML	1.66 (1.29, 2.23)						
LM	(1.30, 2.21)						
CLR	(1.30, 2.20)						
Allele sc.	1.40 (0.99, 1.98)	0.06	33.2	0.007	0.43	NA	4796

IV estimates of the causal assoc. between std. BMD & std. fat mass

Second stage regression



OLS: 1.22 (1.19, 1.26); IV allele score: 1.40 (0.99, 1.98)

Discussion

Multiple instruments:

- Best way to increase precision of IV estimates if can't increase sample size or find stronger IV
- Each instrument should meet with IV assumptions

Discussion

Multiple instruments:

- Best way to increase precision of IV estimates if can't increase sample size or find stronger IV
- Each instrument should meet with IV assumptions
- Investigate joint strength - Cragg-Donald F-statistic
- Investigate over-identification - Sargan test
- Use of allele scores - possible over-identification issue

Discussion

Multiple instruments:

- Best way to increase precision of IV estimates if can't increase sample size or find stronger IV
- Each instrument should meet with IV assumptions
- Investigate joint strength - Cragg-Donald F-statistic
- Investigate over-identification - Sargan test
- Use of allele scores - possible over-identification issue
- This work in:
  Lawlor, Palmer, et al., Statistical Methods in Medical Research, submitted

Acknowledgements

MRC collaborative grant G0601625: Methods for Mendelian randomization

Collaborators: Nuala Sheehan, Vanessa Didelez, Sha Meng, Roger Harbord, John Thompson, Paul Clarke, Frank Windmeijer, Paul Burton, George Davey Smith.

References I

Cragg, J. G., & Donald, S. G. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory*, *9*, 222–240.

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease. *International Journal of Epidemiology*, *32*, 1–22.

Didelez, V., & Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, *16*, 309–330.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., et al. (2007). A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science*, *316*(5826), 889–894.

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, *27*(8), 1133–1163.

Loos, R. J. F., Lindgren, C. M., Li, S., Wheeler, E., Zhao, J. H., Prokopenko, I., et al. (2008). Common variants near mc4r are associated with fat mass, weight and risk of obesity. *Nature Genetics*, *40*(6), 768–775. Available from `http://dx.doi.org/10.1038/ng.140`

Mikusheva, A., & Poi, B. (2006). Tests and confidence sets with correct size when instruments are potentially weak. *The Stata Journal*, *6*(3), 335–347.

References II

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, *20*(4), 518–529.

Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., Mangino, M., et al. (2008). Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, *40*(5), 575–583.

Willer, C. J., Speliotes, E. K., Loos, R. J., Li, S., Lindgren, C. M., Heid, I. M., et al. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics*, *41*, 25–34.