Instrumental variable estimation of the causal risk ratio in cohorts

Tom Palmer Roger Harbord Vanessa Didelez
Nuala Sheehan Debbie Lawlor Jonathan Sterne

MRC Centre for Causal Analyses in Translational Epidemiology,
School of Social and Community Medicine, University of Bristol

August 2010

Outline

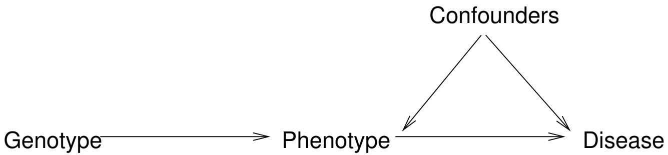
- Instrumental variables
- Causal risk ratio
- Outline estimators
  - apply in ALSPAC BMI-asthma Mendelian randomization example
- Discussion

Instrumental variables (IVs)

X exposure/risk factor/phenotype
Y outcome/disease
Z instrument/genotype

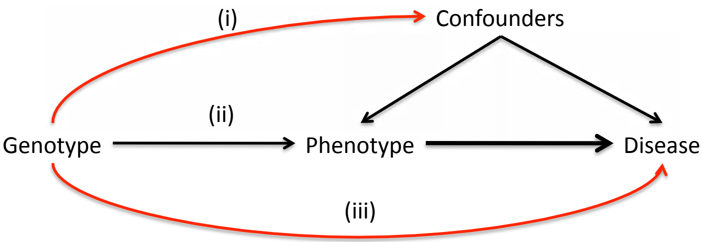
Instrumental variables (IVs)

X exposure/risk factor/phenotype
Y outcome/disease
Z instrument/genotype



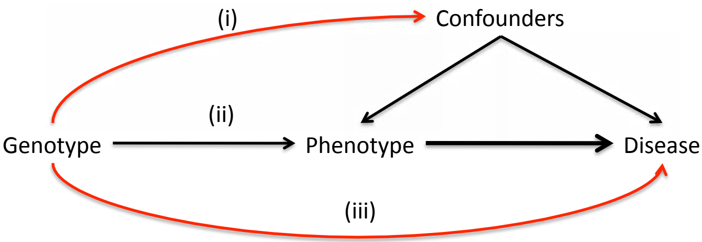
Instrumental variables (IVs)

X exposure/risk factor/phenotype
Y outcome/disease
Z instrument/genotype



Instrumental variables (IVs)

X exposure/risk factor/phenotype
Y outcome/disease
Z instrument/genotype



(i) independent of confounders
(ii) associated with phenotype
(iii) independent of outcome given phenotype and confounders

Causal risk ratio

$$\frac{P(Y = 1|X = x + 1)}{P(Y = 1|X = x)}$$

Causal risk ratio

$$\frac{P(Y=1|X=x+1)}{P(Y=1|X=x)}$$

$$\frac{P(Y=1|do(X=x+1))}{P(Y=1|do(X=x))}$$

Pearl's *do*() operator: set X to x (or potential outcomes)

θ: log CRR

Multiplicative SMM

Multiplicative structural model

Robins, CTSM, 1994; Hernan & Robins, Epi, 2006

Multiplicative SMM

Multiplicative structural model

Robins, CTSM, 1994; Hernan & Robins, Epi, 2006

$$\frac{E(Y=1|X,Z)}{E(Y|X,Z,do(X=0))} = \exp(\theta X)$$

Multiplicative SMM

Multiplicative structural model

Robins, CTSM, 1994; Hernan & Robins, Epi, 2006

$$\frac{E(Y = 1|X, Z)}{E(Y|X, Z, do(X = 0))} = \exp(\theta X)$$

Identification: no effect modification by Z

$exp(\theta)$ causal risk ratio (CRR) in exposed

Multiplicative SMM

Multiplicative structural model

Robins, CTSM, 1994; Hernan & Robins, Epi, 2006

$$\frac{E(Y=1|X,Z)}{E(Y|X,Z,do(X=0))} = \exp(\theta X)$$

Identification: no effect modification by Z

$exp(\theta)$ causal risk ratio (CRR) in exposed

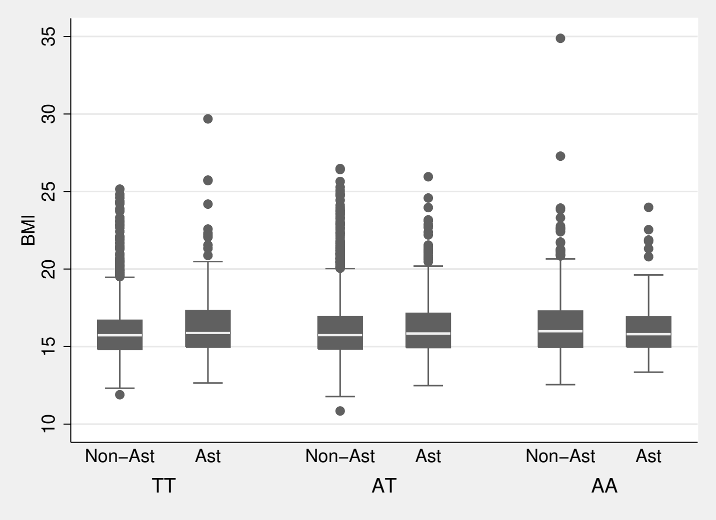
Sample moment condition

$$\sum_i y_i \exp(-\theta x_i)(z_i - \overline{z}) = 0$$

$Y = 1$ provide majority of information

ALSPAC BMI-asthma example

FTO (rs9939609) - BMI - asthma (7yrs)



ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13

* adjusted for sex, birthweight, prenatal maternal smoking, postnatal maternal smoking, maternal education and head of household social class

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)

* adjusted for sex, birthweight, prenatal maternal smoking, postnatal maternal smoking, maternal education and head of household social class

Generalised method of moments estimators

GMM - overidentified models (more Zs than Xs)

Minimise a quadratic form: $Q(\theta) = h'Wh$

Generalised method of moments estimators

GMM - overidentified models (more Zs than Xs)

Minimise a quadratic form: $Q(\theta) = h'Wh$

Form of moment condition $h()$

Generalised method of moments estimators

GMM - overidentified models (more Zs than Xs)

Minimise a quadratic form: $Q(\theta) = h'Wh$

Form of moment condition $h()$

$h() : Z \perp\!\!\!\perp$ function of (Y, X, θ)

Generalised method of moments estimators

GMM - overidentified models (more Zs than Xs)

Minimise a quadratic form: $Q(\theta) = h'Wh$

Form of moment condition $h()$

$h() : Z \perp\!\!\!\perp$ function of (Y, X, θ)

linear: $h() = Z(Y - X\theta)$

First guess at a GMM estimator of the CRR

Additive residual Johnston et al. Stats Med 2007

$$Y = \exp(X\theta) + U$$

First guess at a GMM estimator of the CRR

Additive residual Johnston et al. Stats Med 2007

$$Y = \exp(X\theta) + U$$

if $E(U) = 0$, gives moment condition

$$h() = Z(Y - \exp(X\theta))$$

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54

Additive GMM does not give same estimate as MSMM

Multiplicative GMM

Multiplicative residual

Mullahy, RES, 1997; Windmeijer, JAE, 1997; Angrist, JBES, 2001

$$Y = \exp(X\theta + U)$$

Multiplicative GMM

Multiplicative residual

Mullahy, RES, 1997; Windmeijer, JAE, 1997; Angrist, JBES, 2001

$$Y = \exp(X\theta + U)$$

gives moment condition

$$h() = Z(Y \exp(-X\theta) - 1) = Z\left(\frac{Y - \exp(X\theta)}{\exp(X\theta)}\right)$$

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54
Multiplicative GMM	0.81	0.63, 1.05 (robust SE)

ALSPAC example estimates

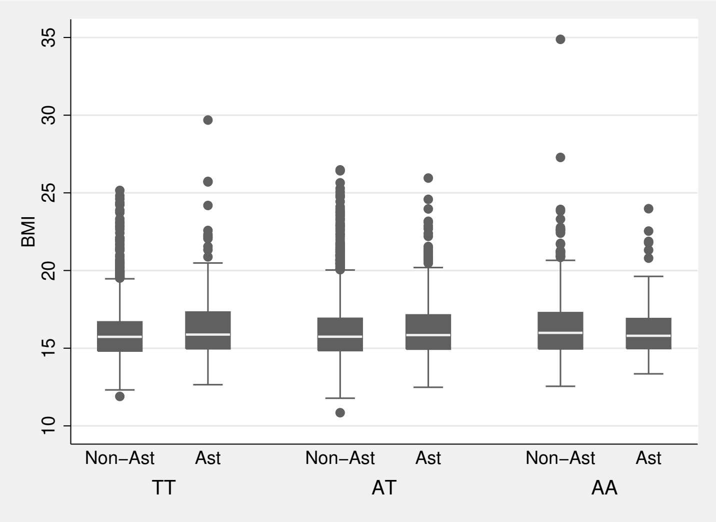
CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54
Multiplicative GMM	0.81	0.63, 1.05 (robust SE)

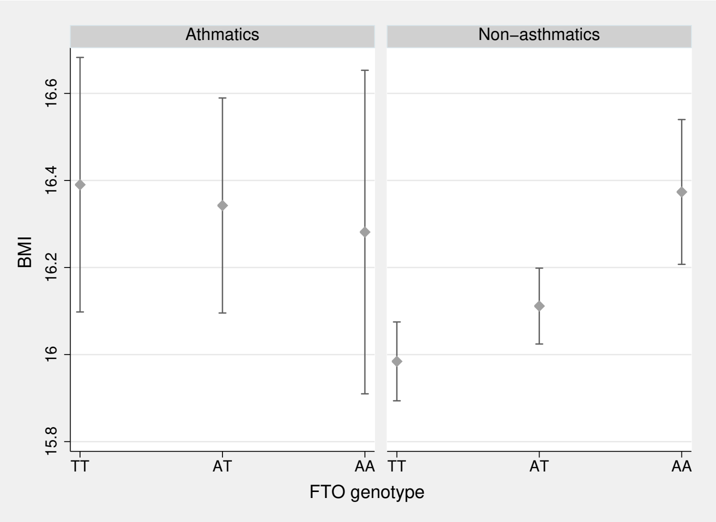
Multiplicative GMM estimate same as MSMM Clarke, Biostats, 2010

User written Stata command `ivpois` Nichols, 2007

MGMM/MSMM other side of null



MGMM/MSMM other side of null



Wald type, two-stage, control function estimators

Wald type estimator: ratio of genotype-disease & genotype-phenotype associations

$$\widehat{\theta} = \frac{\log(RR_{YZ})}{\delta_{XZ}}$$

Wald type, two-stage, control function estimators

Wald type estimator: ratio of genotype-disease & genotype-phenotype associations

$$\widehat{\theta} = \frac{\log(RR_{YZ})}{\delta_{XZ}}$$

Two-stage estimator

Stage 1: linear regression X on Z, save $\widehat{X}$

Stage 2: Poisson regression of Y on $\widehat{X}$

Wald type, two-stage, control function estimators

Wald type estimator: ratio of genotype-disease & genotype-phenotype associations

$$\widehat{\theta} = \frac{\log(RR_{YZ})}{\delta_{XZ}}$$

Two-stage estimator

Stage 1: linear regression X on Z, save $\widehat{X}$

Stage 2: Poisson regression of Y on $\widehat{X}$

Control function estimator Cameron & Trivedi, 2009

Stage 1: additionally save estimated residuals $\widehat{\epsilon}$

Stage 2: Poisson regression of Y on X and $\widehat{\epsilon}$

Rationale: $\widehat{\beta}_{\widehat{\epsilon}} = 0$ endogeneity test

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54
Multiplicative GMM	0.81	0.63, 1.05 (robust SE)

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54
Multiplicative GMM	0.81	0.63, 1.05 (robust SE)
Wald	1.37	0.64, 2.96

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54
Multiplicative GMM	0.81	0.63, 1.05 (robust SE)
Wald	1.37	0.64, 2.96
Two-stage	1.37	0.68, 2.78

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54
Multiplicative GMM	0.81	0.63, 1.05 (robust SE)
Wald	1.37	0.64, 2.96
Two-stage	1.37	0.68, 2.78
Control function	1.37	0.68, 2.76

ALSPAC example estimates

CRR for asthma per unit increase BMI ($N = 4647$)

Estimator	$\widehat{\text{CRR}}$	95% CI
Observational	1.06	1.02, 1.10
Observational*	1.08	1.03, 1.13
Multiplicative SMM	0.81	0.44, 1.48 (bootstrap SE)
Additive GMM	1.21	0.95, 1.54
Multiplicative GMM	0.81	0.63, 1.05 (robust SE)
Wald	1.37	0.64, 2.96
Two-stage	1.37	0.68, 2.78
Control function	1.37	0.68, 2.76

Single IV: Wald, two-stage, control function estimates the same

Multiple IVs: two-stage, control functions estimates the same

Discussion

Additive GMM estimators of CRR & COR do not give same estimates as SMMs

In our example CIs comparable

MSMM/MGMM estimates can be other side of the null

MSMM and MGMM give same estimate of CRR

More technical comparison of estimators Didelez et al., Stats Sci, 2010

Acknowledgements

MRC collaborative grant G0601625

Thanks to Sha Meng, Paul Clarke, Frank Windmeijer, and George Davey Smith.

References I

Angrist, J. D. 2001. Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of Business and Economic Statistics*, **19**(1), 2–28.

Cameron, A. C., & Trivedi, P. K. 2009. *Microeconometrics Using Stata*. College Station, Texas: Stata Press.

Clarke, Paul S., & Windmeijer, F. 2010. Identification of causal effects on binary outcomes using structural mean models. *Biostatistics*. Online: doi:10.1093/biostatistics/kxq024.

Didelez, V., Meng, S., & Sheehan, N. A. 2010. Assumptions of IV Methods for Observational Epidemiology. *Statistical Science*, **25**(1), 22–40.

Hernán, M. A., & Robins, J. M. 2006. Instruments for Causal Inference. An Epidemiologist's Dream? *Epidemiology*, **17**, 360–372.

Johnston, K. M., Gustafson, P., Levy, A. R., & Grootendorst, P. 2008. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statistics in Medicine*, **27**, 1539–1556.

Mullahy, J. 1997. Instrumental–variable estimation of count data models: Applications to models of cigarette smoking behaviour. *The Review of Economics and Statistics*, **79**(4), 568–593.

Nichols, A. 2007. *ivpois: Stata module for IV/GMM Poisson regression*. Statistical Software Components, Boston College Department of Economics. available at http://ideas.repec.org/c/boc/bocode/s456890.html.

Robins, J. M. 1994. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics: Theory and Methods*, **23**(8), 2379–2412.

Windmeijer, F. A. G., & Santos Silva, J. M. C. 1997. Endogeneity in Count Data Models: An Application to Demand for Health Care. *Journal of Applied Econometrics*, **12**(3), 281–294.